UMEÅ UNIVERSITY

In search of future excellence:

Bibliometric indicators, gender differences, and predicting research performance in the early career

Jonas Lindahl

Department of Sociology
Umeå 2020

This work is protected by the Swedish Copyright Legislation (Act 1960:729)
Dissertation for PhD
ISBN print: 978-91-7855-208-5
ISSN: 1104-2508
Akademiska avhandlingar vid sociologiska institutionen,
Umeå universitet
Electronic version available at: http://umu.diva-portal.org/
Printed by: CityPrint i Norr AB
Umeå, Sweden 2020

To Vilja, Herman, Ingeborg, and Malin with love

Table of Contents

Table of Contents ..... **i**

Abstract ..... **ii**

List of original articles ..... **iii**

1. Introduction and aim **1**

The structure of the thesis 4

2. Background: Contextualizing the early career and the disciplines under study **5**

Defining early career researchers 5

Science, technology, engineering, mathematics, and the medical sciences 8

3. Theoretical framework **12**

Theoretical perspectives on the predictability of research performance 12

Cumulative advantage and Merton's reputation-and-resources model *12*

The sacred spark theory *14*

Simonton's model for scientific creativity *14*

Operationalizing research performance with bibliometric indicators 16

What do citations measure: on the validity of operationalizing scientific performance with bibliometric indicators *17*

The correlation between peer review and bibliometric indicators *19*

Bibliometric indicators and research quality *21*

Bibliometric indicators and the concept of research excellence *23*

4. Previous research **25**

Predicting research performance with bibliometric indicators 25

The later career *25*

The early career *27*

The doctoral education *29*

Previous research on gender differences in research performance 31

Gender differences in career development *31*

Gender differences in productivity *33*

5. Results: summary of the papers **38**

Article I 38

Article II 39

Article III 40

Article IV 41

6. Concluding discussion **43**

References **47**

Acknowledgements **60**

Abstract

The governance of higher education institutions and science have endured significant changes during the last decades, emphasizing competitiveness, performance, and excellence. Embedded in this development is an increased use of bibliometric indicators as decision support tools in contexts of e.g., employment, appointment, and funding. These changes have gradually extended to the early career phase and the doctoral education.

The aim of this thesis is to make a contribution to an ongoing discussion about the predictability of research performance and the reasonability of using bibliometric indicators in the early career, with a focus on gender differences. The thesis revolves around three overarching research questions focusing the early career and the doctoral education: (1) the degree to which research performance, as operationalized with bibliometric indicators, is predictable; (2) the degree to which gender differences in early career performance can be explained by research performance during the doctoral education; and (3) to what degree factors such as collaboration and supervisor behaviour, might affect gender differences in research performance.

The main results suggests that research performance in the early career, as operationalized by bibliometric indicators, is predictable. Individuals who publish larger volumes, publish more in high prestige journals, and more excellent research early in their career, are more likely to attain excellence later on. The results also indicates that gender differences in performance can be observed as early as during doctor education and that these differences partly explain the observed performance differences between males and females in the early career. Finally, the results suggests that gender differences in performance during doctoral education can largely be explained by the doctoral student's collaborative networks and supervisor behaviour.

It is concluded that while research performance, as operationalized by bibliometric indicators, during the early career is predictable, there are gender differences in performance that have to be taken into consideration. If they are not, the use of these types of performance indicators in science policy and management might increase the gender gap in science.

List of original articles

I. Lindahl, J., & Danell, R. (2016). The information value of early career productivity in mathematics: a ROC analysis of prediction errors in bibliometricly informed decision making. *Scientometrics, 109*(3), 2241–2262. https://doi.org/10.1007/s11192-016-2097-9

II. Lindahl, J. (2018). Predicting research excellence at the individual level: The importance of publication rate, top journal publications, and top 10% publications in the case of early career mathematicians. *Journal of Informetrics, 12*(2), 518–533. https://doi.org/10.1016/j.joi.2018.04.002

III. Lindahl, J., Colliander, C., & Danell, R., (2020). Early career performance and its correlation with gender and publication output during doctoral education. *Scientometrics, 122*(1), 309–330. https://doi.org/10.1007/s11192-019-03262-1

IV. Lindahl, J., Colliander, C., & Danell, R., (n.d.). The importance of collaboration and supervisor behaviour for gender differences in doctoral student performance and early career development. Unpublished manuscript.

1. Introduction and aim

The governance of higher education institutions and science has endured significant changes in most western countries during the last decades (Whitley & Gläser, 2007). Since the mid-1980s, many countries have implemented national performance-based research funding systems emphasizing competitiveness based on the rationale that research institutions should be rewarded in accordance with their performance (Hicks, 2012). These national systems are a part of the widespread introduction of new public management reforms in western society and in the higher education sector in particular, where the main ideas revolve around increasing productivity and replacing traditional command-and-control systems with market-like incentives, as well as providing a stronger service orientation for research (i.e., aiming for societal value by, for example, serving the high tech economy), and working towards increased autonomy, strategic management, and enhanced accountability (Kettle, 2005).

There are basically two alternatives for assessing research performance in these national performance-based research funding systems: (1) quantitative indicators (e.g., bibliometric indicators that measure publication volume and impact), and (2) peer review (Sivertsen, 2017). While peer review was used in many early performance-based systems, the current trend points towards an increased use of bibliometric indicators (Sivertsen, 2017). However, this trend is not limited to national performance-based research funding systems, and can be viewed as a general trend in science and academia (Hammarfelt, Nelhans, Eklund, & Åström, 2016). Bibliometric indicators are used by universities, institutes, and research councils at all levels (e.g., institutional, departmental, individual) as decision support tools in evaluation and processes of selecting promising candidates for funding, hiring, and appointment (Hammarfelt et al., 2016).[1] This trend has been described as the rise of a "metric culture" in academia (Hammarfelt & de Rijcke, 2015).

[1] For example, many Swedish higher education institutions use some variation of bibliometrics at the individual level to allocate resources internally (e.g., Blekinge Institute of Technology, Karlstad University, Linneaus University, Luleå University of Technology, Södertörn University, the Swedish School of Sport and Health Sciences, and Umeå University). Some of these higher education institutions allocate resources at the individual level throughout the whole institution. In other institutions, the allocation of resources is restricted to specific research areas or faculties. In some of these systems, money is allocated directly to individual researchers as rewards for peer-reviewed articles. At Luleå University of Technology, researchers receive 35,000 SEK (i.e., approximately 3,600 EURO) for an article published in a peer-reviewed journal. If the article is published in a journal that is either indexed in the databases that are accessible through the Web of Science, or a Level 2 journal according to the Norwegian list, the researcher receives an additional 35,000 SEK. Blekinge Institute of Technology also reward individual researchers with an amount of 30,000 SEK for a published journal article indexed in the indices accessible through the Web of Science (Hammarfelt et al., 2016).

One of the main arguments for national governments to implement national performance-based research funding systems is to stimulate the pursuit of research excellence (Hicks, 2012). The excellence argument is often imbedded in a broader discourse about countries' competitiveness in terms of innovation and economic growth in the shift from an industrial-based economy to an information-based economy in a globalized world (Hicks, 2012). Following the performance-based focus of these national systems, new types of funding programmes have been introduced, both governmental and private, designed to identify and target so called "excellent" researchers. The aim of these programmes is to support the most competitive individuals and thus ensure "long-term quality of national R&D efforts and, by extension, secure future economic growth" (Hallonsten & Hugander, 2014).

An implication of the changing political climate in science and academia, emphasizing performance, excellence, and bibliometric indicators, is that the publication track record is increasingly important for the career development of individual researchers. This thesis focuses on two important aspects of this development. First, the importance of excellence and the use bibliometric indicators has gradually extended to the early career phase and doctoral education (McGrail, Rickard, & Jones, 2006; Horta & Santos, 2016). To publish in peer-reviewed journals during doctoral studies is increasingly expected and encouraged (Sinclair, Barnacle, & Cuthbert, 2014). For early career researchers, peer-reviewed publications during doctoral education is viewed as important for gaining an advantage in the competition for funding and employment in the early career phase (Mason, 2018). A sign that the trend of performance and excellence have reached selection processes in the early career is clearly reflected in the development of excellence programmes with a focus on identifying so-called excellent (i.e., the most promising) young researchers and concentrating resources on this top group (Hallonsten & Hugander, 2014).[2]

The second aspect relates to the so called "gender gap" in science (Zuckerman, 2001). The literature on the gender gap in science indicates that: (1) fewer females reach influential positions than their male counterparts (European Commission, 2012); (2) the proportion of females that apply for research funding is lower than the proportion of males (European Commission, 2009; European Commission, 2012); (3) female researchers tend to publish less than their male counterparts (Cole & Zuckerman, 1984; Long, 1992; Xie & Shauman, 1998; Prpić, 2002; Fox,

[2] An example of the use of publication based indicators as decision support tools in the early career phase can be found in the grant "Funding for research time within the field of education science" (i.e., Medel för forskningstid inom det utbildningsvetenskapliga området). This grant is provided by the Umeå school of education at Umeå university in Sweden, and targets early career researchers within the field of education science. Together with the two factors (1) previous external funding, and (2) the research plan, the third factor is a publication-based indicator (Lärarhögksolan, 2019).

2005). Previous research have shown that gender differences in publication output, i.e., a gender gap, can be observed as early as during the doctoral education and the early career (see for example, Epstein, & Lachmann, 2018; Pezzoni, Mairesse, Stephan, & Lane, 2016). The early career phase is often referred to as "the rush hour" of the research career, since it has been shown that events in the early career, for example a rejected application, can have long-lasting consequences for career development (Petersen, Jung, Yang, & Stanley, 2011; Petersen, Riccaboni, Stanley, & Pammolli, 2012; European Commission, 2012). If gender differences in performance during the early career can be explained by performance differences during the doctoral education, the increased use of performance indicators with a focus on identifying excellent individuals may reinforce the observed gender gap in science, especially since it has been shown that small gender differences in the early career tend to increase over time (Zuckerman, 2001).

The basic idea behind the use of performance-based indicators is to enable better decision making (i.e., to distribute resources to the researchers that will make best use of them according to some policy goal). Most previous research has focused on what bibliometric indicators measure, and if bibliometric indicators can be used to evaluate past performance. The question of whether the publication track record of a researcher can be used to predict future performance has gained less attention. Without predictive ability, bibliometric indicators would be useless in the context of selecting individuals for, e.g., employment, appointments, or funding, regardless of how accurately they measure research performance (Danell, 2011). However, even if we can show that we can identify those who will perform well in the future with bibliometric indicators, we also want to make sure that these measures do not reproduce biases or unwanted patterns of inequality between females and males. It is therefore important to understand the degree to which there is predictability and how different indicators affect the predictions. But also if, and if so how, the usage of these indicators may affect the gender gap in science.

The changing political climate with an increased focus on competition, excellence, and bibliometric indicators have resulted in an ongoing discussion in the field of scientometrics about the degree to which research performance is predictable at the individual level as operationalized by bibliometric indicators and the use of these indicators as decision support tools (see, for example, Danell, 2011; Acuna, Allesina, & Kording, 2012; Havemann & Larsen, 2015). This thesis aims to contribute to this discussion by investigating the predictability of research performance as operationalized with bibliometric indicators at the individual level among researchers in their early careers and during their doctoral education with a focus on gender-based differences in performance. The empirical foundation of this thesis is four articles, and the thesis revolves around three

overarching research questions: (1) the degree to which research performance, as operationalized with bibliometric indicators, during the doctoral education and early career is predictable; (2) the degree to which gender differences in early career performance can be explained by research performance during the doctoral education; and (3) to what degree other factors, e.g., collaboration and supervisor behaviour, might affect gender differences in research performance during the doctoral education and the early career. Articles I, II, and partly Article III, are concerned with the first research question. The second and third research questions are examined in Article III and IV.

The structure of the thesis

This thesis is structured as follows: Section 2 present a definition of the early career and a context for the early career researcher. This section also provides a characterization of the research disciplines (i.e., science, technology, engineering, mathematics, and the medical sciences), that have been studied in this thesis. Section 3 presents the main theoretical perspectives of the thesis. Section 4 present previous empirical research and consist of two sub-sections. On the one hand a sub-section that present previous research on predicting research performance with bibliometric indicators. On the other hand a sub-section that present previous research on the gender gap in science and gender differences in research performance. Section 5 consist of summaries of article I, II, III, and VI. The thesis ends with a concluding discussion in section 6.

2. Background: Contextualizing the early career and the disciplines under study

The aim of this section is to present a theoretical framework for defining the early career and early career researchers, as well as characterize the disciplines that have been studied in this thesis and provide motivations for choosing these disciplines.

Defining early career researchers

Defining the early career and early career researchers (ECRs) is important in a thesis focusing on this particular phase in the career. In the context of science policy and grant allocation, the early career is often defined as phase after the doctoral education where researchers is referred to as ECRs (Laudel & Gläser, 2008). The time span for what counts as the early career varies between different actors (i.e., funding bodies, research institutions). However, it is often between the first five and eight years after completion of doctoral education (see e.g., Australian Research Council, 2007; Research Assessment Exercise, 2005; Laudel & Gläser, 2008). Bazeley (2003) provides a definition of the ECR that incorporates the employment situation and a number of other variables:

"*An early career researcher is one who is currently within their first five years of academic or other research-related employment allowing uninterrupted, stable research development following completion of their postgraduate research training*" (Bazeley, 2003, p. 274).

Bazeley's (2003) definition is based on the assumptions that: (1) the completion of doctoral studies is the foundation on which the research career is built; (2) the time period of five years is a sufficient time to start building a track record; and (3) the employment situation is stable enough for the early career researcher to develop a personal research programme (Bazeley, 2003).

Bazeley (2003) provides a reasonable definition of the early career researcher in terms of a timespan in the career where a researcher can be defined as an ECR. However, Bazeley (2003) does not address the complex issues of how the early career differs from the doctoral education and later career phases, and how different aspects of the research career might interact during the early stages of the career.

Laudel and Gläser (2008) propose a theoretical framework suggesting that the academic career consists of three different careers that are interrelated, but to a large extent independent of each other (see Figure 1). These three careers are: (1) the organizational career, which consists of the positions academics attain at different research organizations (e.g., research institutes, universities, companies) during their career; (2) the cognitive career, which consists of the research projects academics are involved in during their career; and (3) the

community career, which consists of the different roles as knowledge producers academics might have in their scientific communities.

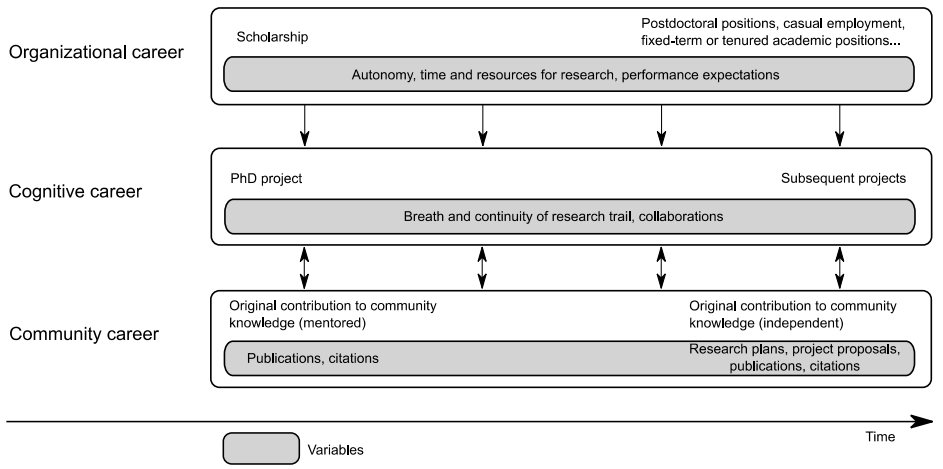

Figure 1. An overview of how the three careers in Laudel and Gläsers' (2008) theoretical framework interact. Adapted from Laudel and Gläser (2008).

The organizational, cognitive, and community careers are interlinked in a complex pattern of interactions (Laudel & Gläser, 2008). For example, the cognitive career (i.e., the research processes the researcher is involved in) is dependent on the community career (i.e., the participation in scientific knowledge production), since the opportunities to conduct research and to collaborate with colleagues and attend conferences partly depend on the "reputation" a researcher has attained in his or her community (Laudel & Gläser, 2008). The cognitive career depends on the organizational career, since it provides the material basis for research, i.e., local work environments, material resources, locally-stored knowledge, research support and colleagues (Laudel & Gläser, 2008). On the one hand, it is within the community career that a researcher builds a strong publication track record and the reputation necessary to be hired by organizations. On the other hand, the organizational career must provide opportunities to participate in scientific knowledge production (i.e., the community career; Laudel & Gläser, 2008).

The early career is, in Laudel and Gläser's (2008) framework, defined by a transitional process where an apprentice is transformed to a colleague. The terms apprentice and colleague are adopted from the four-stage model of the professional career suggested by Dalton, Thompson, & Price (1977), where an apprentice produces research but under the supervision of others who have attained colleague status, and a colleague is an independent researcher who makes original contributions (Laudel & Gläser, 2008). The transition from

apprentice to colleague that occurs in the early career corresponds with the transition from dependent to independent research (Laudel & Gläser, 2008). A colleague, i.e., a full member of the scientific community, is expected to meet the following expectations: (1) be able to evaluate the relevance, validity and reliability of research conducted within his or her field of research; (2) be able to obtain previous research that is relevant to his or her research; (3) be able to identify knowledge gaps and address these gaps with relevant research questions; (4) be able to evaluate whether these research questions can be answered and, if so, answer them in accordance with scientific methods, and communicate the results with the scientific community (Laudel & Gläser, 2008).

The transition from apprentice to colleague is largely a question of attaining autonomy. The transition toward autonomy can be challenging since the knowledge base of ECRs is often relatively small and they may not have formulated a research question or conducted independent research (Laudel & Gläser, 2008). The transition from apprentice to colleague involves all three careers, i.e., the organizational, the cognitive, and the community career (Laudel & Gläser, 2008). The transitional phase in the ECR community career is shaped by expectations to enter the colleague role. With the doctoral thesis, the ECR is expected to have produced a contribution at the "colleague" level, even though with the help of a supervisor (Laudel & Gläser, 2008). After the doctoral education, ECRs are expected to produce contributions at the "colleague" level without the support of a supervisor. The organizational career in the transitional phase consists of the institutional and material conditions to produce research. Becoming an independent researcher, a colleague, is therefore dependent "on the opportunity to make autonomous decisions on research, the time the ECR can spend on research, the resources provided by the organization, and the performance expectations concerning research" (Laudel & Gläser, 2008, p. 392).

The task to determine ECR status is complex. This complexity is illustrated in the three careers model focusing on the transformation from apprentice to colleague, i.e., from dependent to independent researcher, provided by Laudel and Gläser (2008). The early career period might be prolonged if the researcher cannot attain independence for various reasons (e.g., little time for research after doctoral education). As shown by Laudel and Gläser (2008), it is difficult to pinpoint when the transition from apprentice to colleague occurs in the early career. It "can occur in extended PhD phases, postdoctoral positions, other forms of employment, or in the first teaching and research position. Prolonged PhD phases and a series of short-term contracts may either contain the transition period or delay it for several years" (Laudel & Gläser, 2008, p. 401). Similar to Laudel and Gläser (2003), Bazeley's (2003) definition is used in this thesis, where an ECR is a researcher whose career is within five years of completing doctoral education. However, in the two first studies in this thesis (Article I and II), due to

restrictions in the data, the timespan for the dependent variables (i.e., the early career) is four rather than five years. It should also be noted that these two studies use the publication career, i.e., the community career, rather than the organizational or cognitive career, to delineate the early career. The implication of this approach is that the demarcation of doctoral education, the early career, and the later career is less distinct in the first and second articles than in the third and fourth articles.

Science, technology, engineering, mathematics, and the medical sciences

The following sub-section aims to provide a background and some characteristics of the disciplines and research fields that are studied in this thesis. The sub-section also aims to describe the motivation for choosing these disciplines. A first broad delineation could be to categorize these disciplines as belonging to the so-called STEM disciplines (i.e., Science, Technology, Engineering, and Mathematics) and the medical sciences. Articles I and II focus on mathematics, Article III focuses science and the medical sciences, and Article IV includes individuals from the STEM disciplines and the medical sciences.

A basic assumption in this thesis is that scientific disciplines are socially constructed and that disciplines and research fields may differ in terms of knowledge production and scientific communication (Becher & Trowler, 2001), funding and career advancement (Becher & Trowler, 2001), how doctoral education is organized (Hakkarainen, Hytönen, Makkonen, & Lehtinen, 2016), gender equity (European Commission, 2012), etc. The notion of differences between research fields are important in this thesis; there can be differences between research fields that affect the predictability of research performance and gender differences in performance between male and female researchers. In order to characterize the disciplines under study, two theoretical frameworks that focus on differences between disciplines and areas of research are used: Whitely's (1984) characterization of scientific fields as work organizations, and Becher and Trowler's (2001) characterization of scientific fields as academic tribes.

Whitley (1984) aims to provide a framework for understanding research fields as work organizations where scientific knowledge is produced differently depending on context (Whitley, 1984). A central aspect of Whitley (1984) is that research fields are viewed as organizations: "Intellectual fields are seen as the major form of intellectual organizations which structure the framework in which day-to-day decisions, actions, and interpretations are carried out by groups of scientists primarily oriented to public intellectual goals" (Whitley, 2000, p. 8-9).

Whitley's theory is based on two main dimensions, *mutual dependency* and *task uncertainty*, to describe differences between disciplines and research fields. Mutual dependency refers to "scientists' dependence upon particular groups of colleagues to make competent contributions to collective intellectual goals and

acquire prestigious reputations which lead to material rewards" (Whitley, 1984, p. 87). Whitley identifies two types of mutual dependencies: *functional dependency* and *strategic dependency*. Functional dependency refers to the degree to which researchers rely on the results, methods, ideas, and procedures of other researchers in order to make valid knowledge claims. Strategic dependency, on the other hand, reflects how important it is for researchers to persuade their colleagues that their research is important for their field (Whitley, 1984).

The second dimension, *task uncertainty*, depends on the intellectual organization of a research field and refers on the one hand to technical and, on the other hand, to strategic uncertainty. A high technical task uncertainty suggests that conflicts concerning the interpretation of results are common, and that the choice of methods are debated. The degree of strategic task uncertainty depicts the level of consensus on intellectual priorities and the goals of research. Fields with low strategic uncertainty tend to be stable and uniform in terms of the choices of topics and research problems (Whitley, 1984).

Generally, according to Whitley's dimensions, the disciplines that have been investigated in this thesis can be characterized as research fields with a high degree of functional dependence and a low technical task uncertainty. Table 1 represents four categories of fields that share these two dimensions. According to Whitley (1984), such fields can be categorized as: (a) Professional adhocracies; (b) Polycentric professions; (c) Technologically integrated bureaucracies; and (d) Conceptually integrated bureaucracies.

Table 1. Types of scientific fields with a high degree of functional dependence (adapted from Whitley, 1984)

Degree of task uncertainty	Degree of functional dependence: High	
	Degree of strategic dependence	
	Low	High
Low technical and high strategic task uncertainty	(a) *Professional adhocracy* producing empirical, specific knowledge, e.g., bio-medical science, artificial intelligence, engineering, pre-Darwinian nineteenth century ornithology	(b) *Polycentric profession* producing specific, theoretically co-ordinated knowledge, e.g., experimental physiology, continental mathematics
Low technical and low strategic task uncertainty	(c) *Technologically integrated bureaucracy* producing empirical, specific knowledge, e.g., twentieth-century chemistry	(d) *Conceptually integrated bureaucracy* producing specific, theoretically oriented knowledge, e.g., post-1945 physics

Another perspective on the question of how to characterize differences between disciplines in research is suggested by Becher and Trowler (2001), who identified "four basic sets of properties: hard/soft and pure/applied in the cognitive realm;

convergent/divergent and urban/rural in the social" (Becher & Trowler, 2001, p. 184). The categorization according to the cognitive realm gives rise to four disciplinary categories. The hard/pure category contains disciplines from the natural sciences (e.g., physics, chemistry, biology) and mathematics. The hard/applied contains disciplines related to the science-based professions, e.g., engineering fields and computer science. The soft/pure and soft/applied categories encompass disciplines related to the social professions (e.g., education, social work, law), and the humanities and social sciences (Becher & Trowler, 2001). The disciplines included in this thesis (i.e., STEM and the medical sciences) belong in the hard/pure and hard/applied categories and are characterized by more or less unified paradigms where the knowledge production is cumulative (Becher & Trowler, 2001).

In the social realm, the convergent/divergent dichotomy has to do with the degree to which standards, perspectives, procedures, and problems are agreed upon. A typical example of a convergent discipline is physics, where there is "an overriding sense of collective kinship, a mutuality of interests, a shared intellectual style, a consensual understanding of 'profound simplicities', and even 'a quasi-religious belief in the unity of nature'" (Becher & Trowler, 2001, p. 187). The urban/rural dichotomy reflects the researcher-to-problem ratio, where a high ratio indicates high competition for positions and resources, as well as competition for solving contemporary problems before others in the community. According to Becher and Towler (2001), physics and biochemistry can be considered examples of urban disciplines. Chemistry, on the other hand, is categorized as rural since it is "uncommon for researchers to compete in 'a fight for the finish'; publication delays can be quite lengthy in most subdisciplines, and the circulation of preprints is noticeably less prevalent than in physics; laboratory investigation 'doesn't involve expensive machines, experiments on a large scale, or collaboration in a global sense" (Becher & Trowler, 2001, p. 189).

To summarize, the disciplines and research fields included in this thesis (i.e., science, technology, engineering, mathematics, and the medical sciences) can, according to the frameworks of Whitley (1984) and Becher and Trowler (2001), generally be characterized as hard/pure and hard/applied, and as fields with a high degree of functional dependence and a low technical uncertainty. A common denominator and motivation for choosing these types of disciplines in this thesis is that the main channel for scientific communication is generally scientific journals and peer-reviewed articles. Scientific communication in these disciplines is therefore well covered in the large citation databases (e.g., SCI-EXPANDED, SSCI, A&HCI, accessible through the Web of Science; Moed, 2005), which is a requirement for collecting citation data and constructing citation-

based bibliometric indicators.[3] Thus, one of the motivations for choosing STEM disciplines and the medical sciences is data driven.

Another important distinction that has guided the choice of disciplines and data collection in this thesis lies between research areas with a high degree of collaboration, e.g., chemistry, physics, bio-medical science (as represented in Articles III and IV), and areas with a lower degree of collaboration, e.g., mathematics (as represented in Articles I and II). Mathematics differs from the other STEM disciplines on two points: (1) mathematicians usually write articles as single authors or in small teams, i.e., the degree of collaboration is generally low; and (2) external resources are less important for conducting research and producing scientific knowledge, which is often related to the degree of collaboration in a field. The choice of mathematics enables an examination of a discipline where knowledge production is individually driven by researchers who are less dependent on resources and collaboration as opposed to the more collaborative disciplines.

[3] An exception here is mathematics, that does not have good coverage in the large citation databases. These databases was therefore not used to collect the data for Article I and II (where mathematics is the main discipline under study).

3. Theoretical framework

This section starts with a sub-section where theoretical perspectives on the predictability of research performance are presented. The section ends with a sub-section on citation theory and bibliometric indicators.

Theoretical perspectives on the predictability of research performance

What are the theoretical bases for assuming that research performance is predictable over time? This question is important in the context of this thesis. The purpose of this section is to present the main theories that have been addressed in previous literature to enable investigations on questions related to issue of whether, and if so to what degree, research performance is characterized by continuity, and thus predictability. It should be underlined that the aim of this thesis is not to test specific theories or mechanisms, or how much variability a specific theory or mechanism explains. Rather, the theories are used to answer the question why it is reasonable to assume that it is possible to make these types of predictions in the first place. In the literature there are two main theories. Theories of cumulative advantage and the sacred spark theory, suggesting that the process of scientific performance is, in terms of productivity and impact, predictable. There are also theories suggesting that the process of scientific performance is characterized by a significant degree of randomness and unpredictability (De Bellis, 2009; DiPrete & Eirich, 2006). In this section, the main contributions from each of these positions are presented. Cumulative advantages in science and Merton's reputation-and-resources model as an explanatory mechanism is presented first (Merton, 1973; Merton, 1988; Merton, 1968), followed by the sacred spark theory (Cole & Cole, 1973; Rodgers & Rodgers, 1999), and finally randomness in scientific performance and Simonton's model of scientific creativity, where a restricted amount of chance and unpredictability is assumed to be a central feature in processes of scientific performance (Simonton, 2003; Simonton, 2004).

Cumulative advantage and Merton's reputation-and-resources model

In processes of cumulative advantage, small advantages accumulate over time and produce patterns of inequality between researchers (DiPrete & Eirich, 2006). The logic of cumulative advantage suggests that the occurrence of an event increases the probability of the recurrence of such an event in the future (Price, 1976). In the context of science and academia the achievement of getting an article published is rewarded (by some mechanism) with an increased probability

of getting future articles published. Such logic implies continuity between past and future research achievements.

According to Merton (1973), science can be understood as a social institution governed by social norms. The four main norms are: (a) universalism, i.e., that science should be impartial and unbiased; (b) communism, i.e., that the results of science should be freely available to the scientific community; (c) disinterestedness, i.e., that those involved in science should work for the common good of the scientific enterprise and not for other gains; and (d) organized scepticism, i.e., avoidance of dogmas etc. The existence of these norms is explained by the function they have in reaching the goals of science as a social institution, which is to produce new scientific knowledge. In addition to the normative structure of science, Merton developed a theory of why researchers would want to pursue the main goal of science. This theory is often called the reward system of science, according to which it is functional for the goal of science to develop a structure that reward researchers who produce new scientific knowledge.

In Merton's reputation-and-resources model, processes of cumulative advantage in science are – according to DiPrete and Eirich (2006) – explained on the basis of three premises that relates to the reward system of science, (1) researchers compete for a limited amount of resources in science, (2) it is difficult to observe and measure scientific talent directly, and (3) resources in science are distributed with respect to the Mertonian norms of (a) universalism, by which recognition in science should be based only on quality, not on other factors such as gender, race, nationality, religion, class, personal traits, etc., and (b) communism, by which "resources should be allocated to maximize the overall productivity of the scientific community" (DiPrete & Eirich, 2006, p. 281). The purpose of resource allocation in this model is not to reward past performance, but rather to enable future research achievements (DiPrete & Eirich, 2006). Following the logic of cumulative advantage, the model proposes that past scientific success enhances the probability of getting resources, and resources increase the probability of producing quality research, which subsequently enhances the probability of more resources, and so on (Merton, 1988).

The main consequences of the reputation-and-resources model according to DiPrete and Eirich (2006) is that (1) the difference in rewards between the higher performing and lower performing researchers would increase with time (DiPrete & Eirich, 2006), (2) chance events, e.g., a grant early in the career, could lead to advantages for one out of two researchers with zero difference in talent and other relevant preconditions (Cole, Cole, & Simon, 1981), and (3) according to the so-called Matthew effect – which refers to a tendency in science where eminent researchers get more credit than relatively unknown researchers for comparable

scientific contributions (Merton, 1968) – researchers with great reputation and visibility in science attract more rewards than unknown researchers for scientific work of comparable quantity and quality (DiPrete & Eirich, 2006). An effect that is related to the Matthew effect and relevant for this thesis is the so-called Matilda effect in science (Rossiter, 1993). The Matilda effect suggests that female contributions to science are systematically undervalued compared with male contributions of comparable quality (Rossiter, 1993).

Merton's normative view of science, the reputation-and-resource model and the theory of cumulative advantage suggest that researchers' performance is predictable over time. Furthermore, the basic assumptions made by the reputation-and-resources model seem to have been adopted in contemporary science policy and management (Whitley, 2007).

The sacred spark theory

The sacred spark theory is positioned within literature that focuses on individual-level variables and psychological, e.g., cognitive and emotional, traits of scientists to explain performance differences between researchers (Fox, 1983). Cole and Cole (1973) first coined the term to describe successful researchers that seemed to be motivated by a sacred spark. The basic assumption in the sacred spark theory is that differences in scientific performance among researchers can be explained by "substantial, predetermined differences among scientists in their ability and motivation to do creative research" (Allison & Stewart, 1974, p. 596).

The sacred spark theory suggest, as does the reputation-and-resource model, that researchers' performance is predictable over time. However, while the sacred spark theory relates the predictability of research performance to inherent traits that are stable over time, e.g., a researcher who performs well in their early career is likely to continue to perform well in their later career because of an inherently high ability to do creative research, the reputation-and-resource model, and the theory of cumulative advantage suggest that early performance actually affects later performance since early high performance leads to more resources that enable later success (Bornmann, & Williams, 2017b).

Simonton's model for scientific creativity

In contrast to models for scientific performance based on the theory of cumulative advantage and the sacred spark theory, Simonton's psychologically oriented model for scientific creativity suggests that the process of scientific performance "demands the intrusion of a restricted amount of chance, randomness, or unpredictability" (Simonton, 2003, p. 476). Simonton assumes a combinatorial process underlying scientific creativity. Researchers within a particular scientific domain have a restricted set of facts, concepts, techniques,

heuristics, themes, questions, goals, criteria, etc. This population of ideas constitutes the ideational foundation of the domain. The actual process of scientific creativity consists of finding useful combinations of ideas given personal and domain-specific restrictions. Such combinations potentially lead to scientific publications (Simonton, 2003). This model is set within a Darwinian framework where the process of scientific performance is viewed as a blind variation-selection process (Simonton, 1997). Researchers in every scientific domain produce a number of ideas and publications. Concerning these ideas and publications, the variations-selection process operates at three levels: (1) the individual cognitive (i.e., the individual selection among ideational combinations); (2) the interpersonal (i.e., competition between manuscripts to get published); and (3) the socio-cultural (i.e., the long-term impact of a publication in the scientific community). Since the selection process is active at different levels, the selection at the individual level is blind to the final reception of the work at the interpersonal and socio-cultural levels. Likewise, the selection at the interpersonal level is blind to the long-term impact of the socio-cultural level due to, for example, changes in norms and values, the emergence of new technologies or new knowledge (Simonton, 1997).

Given that the variation-selection process of scientific creativity is blind, it is hypothesized that combinations of high and low quality are randomly distributed across (i.e., cross-sectional perspective) and within (i.e., longitudinal perspective) researchers' careers (Simonton, 2003; Simonton, 2004). To be more specific, the two main hypotheses suggested by Simonton's model for scientific creativity are:

(1) Due to the low probability of finding useful combinations, the distribution of publications in any given year follows the Poisson distribution. The Poisson distribution describes the probability of the occurrence of an event within a given unit of time and applies when the probability of the actual occurrence of an event is very low (Simonton, 2004).

(2) That the relationship between quality and quantity in science is best explained by a so-called Constant-probability-of-success model, which basically states that quality (i.e., number of citations) is a probabilistic consequence of quantity (i.e., number of publications), e.g., the most productive periods in a researcher's career are also the periods where most high quality papers are produced (Simonton, 1985; Simonton, 1997; Simonton, 2004). That scientific performance is a process characterized by randomness also has support in the bibliometric literature. In a series of articles Huber and Wagner-Döbler (Huber, 1998, 2001; Huber & Wagner-Döbler, 2001a; Huber & Wagner-Döbler, 2001b) showed that researchers – contrary to the assumptions held by cumulative advantage – publish at a constant rate during their careers with random fluctuations, fitting the Poisson distribution.

While acknowledging randomness and chance in scientific creativity, the constant-probability-of-success model suggests that the mechanism that would explain continuity in performance (i.e., producing high quality research) over time, is a researchers ability to produce large quantities of scientific output (for example articles or books).

The guiding theory for this thesis is cumulative advantages and the reputation-and-resource model. This theoretical framework works well with the characteristics of the disciplines and research fields that are studied in this thesis. Fields with a high degree of functional dependency and a low technical task uncertainty according to Whitley (1984), that align to the hard/soft and hard/applied categories according to Becher and Trowler (2001), for example, chemistry, biology, physics, engineering, and the medical sciences, are often very resource dependent and collaborative. A limitation with the sacred spark theory and Simonton's model for scientific creativity is their focus on the individual and that they focus less on the collaborative and social dimension of science. Scientific knowledge production is for the most part a collaborative enterprise. However, it should be acknowledged that these theories are probably not mutually exclusive. For example, the degree of sacred spark among researchers (e.g., how creative, intelligent, motivated, and/or persistent the individual is; Rodgers & Rodgers, 1999), may influence productivity and contribute to processes of cumulative advantage. It should also be noted that the explanatory mechanism might be different in different fields of research (Danell, 2011). It could be the case that aspects of the two more individually oriented theories better explain the predictability of research performance in a discipline such as mathematics, where knowledge production is more individually driven.

Operationalizing research performance with bibliometric indicators

The predictability of future research performance based on bibliometric indicators depends on two things: (1) the extent to which research performance is accurately measured by bibliometric indicators, and (2) the extent to which performance is correlated over time. In the previous section we were interested in the theoretical basis for the second point. In this section we are interested in the first point, i.e., the validity of using bibliometric indicators to measure research performance, and research quality and quantity. The first sub-section presents an overview of citation theory. The second sub-section is concerned with the relationship between bibliometric indicators and peer-review assessments. In the third sub-section, the relationship between bibliometric indicators and the concept of research quality is presented.

What do citations measure: on the validity of operationalizing scientific performance with bibliometric indicators

The question of what citations measure has been an important one in scientometric research for a long time (Aksnes, Langfeldt, & Wouters, 2019), because the legitimacy of using citation-based indicators depends on their validity as measures to assess the value of scientific contributions (Gilbert, 1977).

Merton's normative theory of the reward system of science is "often considered to have provided the original theoretical basis for linking citation counts to the use and quality of scientific contributions" (Aksnes et al., 2019, p. 5). The most important form of reward in this system is peer recognition for scientific performance. The act of citing another researcher's work is an act of giving recognition to previous research in accordance with the Mertonian norms, i.e., researchers pay intellectual debts by referencing research that has been influential to them (Small, 2004). For simplicity we may refer to this view of citations as the normative position. Thus, the normative position suggests that researchers reward other researchers whose work has been important in their own research by citing these works. Researchers give credit where credit is due (Small, 2004). As such, citations are interpreted as a measure of intellectual influence (Zuckerman, 1987). Merton's normative theory has generated a number of interpretations of the citation that are related by founding their interpretation on the normative view of science (Wouter, 1999). However, a common denominator of these interpretations is that the number of citations can be regarded as a valid measure of the impact or influence of a publication or units at aggregated levels, e.g., research teams, departments, or institutions (Aksnes et al., 2019).

The most common critique of interpretations and theories of citations based on Merton's normative system of science can be found in different studies focusing on the motives behind the practice of giving references. It has been shown that reference behaviour depends on many different motives, and might have many different functions (e.g., "providing background reading, identifying methodology, paying homage to pioneers, identifying original publication or other work describing an eponymic concept, identifying an original publication in which an idea or concept was discussed, giving credit for related work, criticizing previous work, correcting a work, substantiating claims, alerts to a forthcoming work, providing leads to poorly disseminated work, authenticating data and classes of fact, disclaiming works of others, and disputing priority claims" Aksnes et al., 2019, p. 4), often contrary to the Mertonian norms of science. For simplicity, we refer to this critique as the social constructivist position. The social constructivist position generally interprets the heterogeneity of referencing behaviour as an indication that variations in citation counts

basically reflect “an aggregate of a wide variety of individual motives and special circumstanneces” (Moed, 2005, p. 213) determined by social context.

One strand of work in the social constructivist position suggest that the act of referencing is a rhetorical device used by authors for the purpose of persuading the audience of the significance, validity, and novelty of their research (see e.g., Latour, 1987). From the social constructivist perspective, citing is not an act of repaying intellectual debts and not an activity in line with the social norms of science, thus undermining the use of citations as a measure of intellectual influence, as suggested by the normative position. However, in this position we also find works pointing toward risks with biases in citation measures based on factors regulating the potential to receive citations regardless of the content and quality of the actual publications, e.g., language (i.e., the citation bases used to collect data consisted mostly of research in English); the size of research field (i.e., the size of the audience; small research fields have fewer potential citers and therefore receive fewer citations compared to larger fields); country of origin; degree of specialization; access for the intended audience, strategic overciting, etc (see e.g., Zuckerman, 1987; Kärki, Kortelainen, Eriksson, & Ungern-Sternberg, 1998).

The social constructivist position views the heterogeneity of referencing behaviour as evidence that citations are not given on the basis of the Mertonian norms, but on the basis of a variety of individual motives, which may lead to unwanted biases if citations are used in contexts of, for example, research evaluations. There seems to be some consensus in the scientometric community that referencing behaviour is biased to some degree. However, “it has been argued that this bias is not fatal for the use of citation as performance indicators – to a certain extent, the biases are averaged out at aggregated levels” (Aksnes et al., 2019, p. 5). For example, van Raan (1998) argues that we should look at citation aggregates, rather than individual reference behaviour when we are determining the usefulness of citations, and suggests that we should separate reference analysis, where there is one citing document and many cited documents, from citation analysis, where there are many citing documents and one cited document. The reference list in a particular paper may have problems such as missing references to significant papers, missing references to seminal authors in the specific field of research, or an exaggerated amount of references to some specific author. van Raan asks the following question concerning the observed heterogeneity in referencing behaviour: “Does this mean that one would never be able to get any sensible idea of the most important work in that field?” (van Raan, 1998, p. 134). His conclusion is no, since this would be the case only if all researchers compiled their reference lists completely at random, and such a claim is not supported by empirical research (van Raan, 1998). van Raan (1998) concludes that as long as reference behaviour is not a completely random process,

valid patterns can be extracted from citation data with respect to scientific influence at different levels of aggregation, and claims that: “there is no sound empirical evidence that citation-biases are the predominant character of reference lists in scientific papers, and that less predominant citation-biases do not cancel each other” (van Raan, 1998, p. 135). Luukkonen (1990) suggests that there are differences in the meaning of citations, and that the motives for citing do not invalidate citation-based indicators as measures of research performance, since motives for citing and the consequences of being cited are analytically distinct.

In the late 1990s, the opposition between the normative theories of citations and the social constructivist theories of reference behaviour seemed to have resulted in something like a deadlock (Wouters, 1999). Since the late 1990s, there have been some attempts in the literature to find a theoretical framework that could resolve this deadlock. Most of these attempts focus on finding a framework that can somehow fit both the normative and social constructivist positions under one umbrella, see e.g., Amsterdamska & Leydesdorff’s (1989) multidimensional approach; Cozzens’ (1989) pluralistic explanation of citations; Wouters’ (1999) reflexive indicator theory; Small’s (2004) citation cube model; Aksnes’ (2003) conceptual distinction between quality dynamics and visibility dynamics in reference behaviour and citation aggregates; and Riviera’s (2013) view of scientific communities as autopoietic systems. Currently, the theoretical basis for citations in terms of a unifying theory can be considered an open problem in scientometrics. However, in practice the normative position has largely been adapted as the theoretical foundation of citations by the scientometric community and contemporary science policy and management (Whitley, 2007; Moed, 2005).

The correlation between peer review and bibliometric indicators

Previous research has shown that there are many factors at work in the citation process. However, the validity of citation-based indicators has also been tested by comparing citation-based indicators with the outcomes of peer review (Aksnes et al., 2019). Traditionally, peer review has been used as the primary method to assess research performance (Abramo, D'Angelo, & Reale, 2019). Peer review assessments have therefore been considered a baseline against which citation-based indicators can be validated. Since peer review is accepted as a valid method to evaluate research, it is assumed that if the outcomes of peer review and citation-based indicators are correlated this would indicate that the citation-based indicators can be used as valid indicators of scientific performance (Aksnes et al., 2019).

During the last decades, numerous studies have investigated the correlation between peer review and citation-based indicators. It should be mentioned that the studies are somewhat heterogeneous in terms of methodology and level of enquiry. One line of research focuses on the correlation between citation rates of individual works and peer review (see e.g., Bertocchi, Gambardella, Jappelli, Nappi, & Peracchi, 2015; Bornmann & Leydesdorff, 2013; Allen, Jones, Dolby, Lynn, & Walport, 2009; Reale, Barbara, & Costantini, 2007; van Raan, 2006; Aksnes & Taxt, 2004; Oppenheim & Norris, 2003). Another line of research focuses on researchers at the individual level by analyzing grant peer review decisions and to what degree applicants that have been awarded funding are more cited than unfunded applicants (see e.g., Cabezas-Clavijo, Robinson-García, Escabias, & Jiménez-Contreras, 2013; Hornbostel, Bohmer, Klingsporn, Neufeld, & Von Ins, 2009). At higher levels of aggregation there are lines of research focusing on the correlation between peer review and bibliometric indicators at the level of: (1) research groups (see e.g., Rinia, van Leeuwen, van Vuren, & van Raan, 1998; Aksnes & Taxt, 2004; van Raan, 2006); (2) departments at research institutions (see e.g., Oppenheim, 1997); and (3) research institutions (see e.g., Pride & Knoth, 2018; Abramo, D'Angelo, & Di Costa, 2011; Franschet & Constantin, 2011; Thomas & Watkin, 1998).

Most of these studies have found a positive correlation between peer review and citation-based indicators, and there seems to be consensus in the literature that there is a positive statistical relationship between peer review and citation-based indicators (see e.g., Aksnes, Langfeldt, & Wouters, 2019; Abramo et al., 2019; Colliander, 2014). However, a number of problems with these studies have been identified that present some limitations to what conclusions can be drawn from them: (1) the studies use different methods, different bibliometric indicators and different ways to construct these indicators. They also cover different research areas and scientific sectors. This makes generalization of the results difficult (Abramo et al., 2019; Colliander, 2014); (2) the assumption that peer review should be the "ground truth" for validating bibliometric indicators may be problematic. For example, the evaluation criteria in peer review may consist of factors besides scientific quality or aspects of scientific quality that are not likely to be covered by citation-based indicators. Peer review rating of quality may consist of many different dimensions besides impact, e.g., originality, significance, rigour, vitality, sustainability (Aksnes et al., 2019; Colliander, 2014). Further, the reliability of peer review may be questionable due to peer biases, mistakes, or lack of competence, and it has been shown that pure chance might play quite a large role in peer review outcomes (Aksnes et al.,2019; Rothwell & Martyn, 2000; Nederhof, 1988); and (3) during the last decades it has become increasingly common for peer review panels to use bibliometric indicators in the peer review evaluation procedures, which makes the independence between the two methods questionable (Aksnes et al., 2019).

Bibliometric indicators and research quality

Citations are often explicitly or implicitly associated with the concept of research quality. Research quality is a multidimensional concept and, in this section, the complex relationship between citation-based indicators and research quality is reviewed. The goal of this section is to find out what aspects of research quality, if any, citation indicators may represent.

In their review, Aksnes et al. (2019), distinguish four key dimensions of research quality: (1) plausibility; (2) originality; (3) scientific value; and (4) societal value. In this sub-section section the relationship between these dimensions and citation-based indicators is discussed on the basis of Aksnes et al. (2019).

The first key dimension of the concept of research quality revolves around the plausibility, soundness, and solidity of research. These aspects also include scientific virtues, e.g., that research should be well-founded, based on scientific methods, and produce convincing results (Aksnes et al., 2019). Few empirical studies have investigated how citations relate to or reflect these aspects of the concept of research quality. Although some studies have shown that publications that have been retracted due to fabrication and falsification can be highly cited, which indicates that these aspects do not necessarily equate to high citation counts. However, this category of publications is a very narrow object of analysis. Aksnes et al. (2019) suggest that potential citers, i.e., researchers writing articles, avoid citing articles that have poor solidity or plausibility and tend to cite those with high solidity and plausibility. If this line of reasoning is correct, it could be the case that solidity/plausibility as perceived by the citing researcher is reflected in citation patterns (Aksnes et al. 2019). This issue has not been addressed much in the literature. However, given that it has been shown that many factors not associated with the solidity of research products are involved in the citation process, Aksnes et al. (2019) conclude that "it seems unlikely that citations can be seen as valid indicators of the solidity of the publications" (p. 9).

The second key dimension of the concept of research quality, originality or novelty, is related to the fact that research, in order to live up to scientific standards, should produce new knowledge in terms of, e.g., new hypotheses, new methods, new theories, new empirical results (Aksnes et al. 2019). Some results in the literature seem to support the hypothesis that research with high originality also tends to receive high citation counts. One example of this is that Nobel Laureates tend to attain higher citation counts than average (Gingras & Wallace, 2010; Wagner, Horlings, Whetsell, Mattsson, & Nordqvist, 2015; Garfield & Welljams-Dorof, 1992). However, it has also been shown that high citation counts might not necessarily indicate breakthrough or Nobel-class research (Aksnes et al. 2019). In their review, Aksnes et al. (2019) conclude that it is uncertain to what degree the dimension of originality and citation counts are related, since studies

with high originality may receive few citations due to, for example, the fact that the research question is a "dead end" and not a meaningful line of enquiry for future research.

The third key dimension, scientific value or significance, is related to the relevance and importance of a work to previous and future research, and may depend on, for example, generalizability, size, and interest in the research problem/question (Aksnes et al. 2019). Arguably, scientific value or significance is the dimension that citation counts are most directly related to (Aksnes et al. 2019). When a researcher cites another publication, it is assumed that the cited publication has been useful or relevant for conducting the research or writing the paper. It follows that papers that receive more citations have been more useful than papers with fewer citations (Aksnes, 2005). In line with the normative view of citations, this means that citation counts can be understood as a measure of a publication's usefulness, impact, or influence on following research (Aksnes et al. 2019). However, it should be noted that there is a conceptual difference between quality and impact. Quality can be viewed as an inherent property of a publication describing, for example, how well structured the work is, if it is free from errors, or the degree to which the work contribute with original scientific knowledge (Martin & Irvin, 1983). Impact, on the other hand, can be defined as the degree to which a publication influences research at given point in time. Aksnes et al. (2019) conclude that impact is the dimension of research quality that best describes what citations measure, since, for example, a "mistaken" publication can receive high citation counts, and a more recognized researcher might receive higher citation counts only due to higher visibility in the community.

The fourth key dimension, societal value, refers to any value external to the scientific value that is relevant for society in some way, e.g., for education, health, the economy, or the environment. There seems to be consensus in the scientometric community that traditional citation-based indicators cannot adequately measure societal value. Even though funding agencies often take societal value into consideration as a review criteria (Langfeldt & Scordato, 2015), and the question of developing indicators and methods to measure societal value is an active problem area, the question of measuring societal value must still be considered an unresolved problem (Aksnes et al. 2019).

To summarize, while there is no evidence, or too little empirical evidence, that citations reflect the dimensions of plausibility, originality, or societal value, it seems reasonable to assume that they reflect the key dimensions of scientific value, i.e., the relevance and importance of the research under assessment for future research (Aksnes et al. 2019). This aspect of what citation-based indicators measure should be taken into consideration when interpreting the results of this thesis.

Bibliometric indicators and the concept of research excellence

The concept of research excellence have, with the changing political climate in science and academia, been adopted as a policy priority in many countries worldwide. However, the understanding of the concept and how it is integrated in research policy, research management practices, and decision-making, varies to a large extent between countries (Tijssen, 2003). It is difficult to find a clear and consensual definition of what constitutes research excellence (Addis & Pagnini, 2010). In this section we will look at some suggestions on how to understand and define the concept; how it is related to bibliometric indicators and how the concept is understood and used in this thesis.

Addis and Pagnini (2010) suggest the following definition:

"Scientific excellence is the ability of a scientist or an institution to impact on a field of study producing a major change, leading other scientists towards asking new questions and producing new, important and useful contributions to knowledge, using new methodologies. The quality of excellence must be proven by a number of means, (such as publications, citations, funding, and students) and recognized by peers by the bestowing of various honours, prizes and other awards" (Addis & Pagnini, 2010, p. 7).

Two central aspects of the concept can be deduced: (1) the concept of research excellence is related to the concept of quality. In the context of research evaluation, Hellström (2011) suggest that excellence can be understood as a measure of high, to very high, scientific quality; and (2) similar to the concept of quality, scientific excellence can be considered a multidimensional concept relating to dimensions such as impact, recognition, originality, and leadership (Addis & Pagnini, 2010).

While excellence is related to quality and can be considered a multidimensional concept, a third important characteristic of research excellence, as suggested by Tijssen (2003), has to do with it being a comparative and expressing superiority in terms of scientific quality and/or quantity, rather than denoting an inherent quality. This aspect of excellence is important in the context of this thesis since (1) it adhere to the analytical approach of benchmarking where the units under evaluation is compared with some standard, and (2) the operationalization of research excellence with bibliometric indicators is often based on such a view of the concept. In practice research excellence is often associated with some specific level of quantity or quality of research output, rather than explicitly defined (Addis & Pagnini, 2010). For example, the European Research Councils 2010 Working Programme had several criterion for applicants that adhere to specific benchmarks, for example, 10 publications in major international peer-reviewed scientific journals (Addis & Pagnini, 2010).

In this thesis, the concept of research excellence is understood measure of high, to very high, scientific quality and it is operationalized by means of bibliometric indicators. Since the end of the 1990s the scientometric community have focused on highly cited publications in the upper end of the citation distribution to operationalize research excellence (Bornmann, 2014). This approach to identify excellence is based on the observation that citation distributions are very skewed. At the level of individual articles there are a few articles that are highly cited and many articles with few or no citations at all (Bornmann, 2014). It should be noted that since research excellence is a multidimensional concept the operationalization of excellence with bibliometric indicators will not, similar to the concept of scientific quality, be able to account for all aspects of excellence.

The recommended best practice for operationalizing the concept of excellence is to use relative percentile-based citation indicators that define highly cited papers as having a citation rate above some citation threshold (Bornmann, 2014). While there is a best practice concerning how to calculate these types of indicators, there are no general consensus on how to set the thresholds for excellence. Can an article be considered excellent if it have received a citation count equal to or above the 10th, 5th, 1st percentile in the citation distribution of set of comparable articles? In an empirical study, Bornmann (2014) examines the most commonly used operationalizations of research excellence in 305 research articles and found that in the group of articles that used operationalizations of excellence in accordance with best practice, the top 1% and the top 10% was the two most commonly used thresholds. While there is a recommended best practice in the literature on how to operationalize research excellence at the level of publications, it is more difficult to find a consensual best practice in the literature on how to set the thresholds when operationalizing research excellence at the individual level (Costas & Noyons, 2013). Costas and Noyons (2013) suggest that a so called "top topper" (i.e., an excellent researcher) is an individual that is within the 25th percentile in terms of productivity, average impact, and journal impact. Another approach is to provide a definition of excellent publications and in turn define authors of such publications as excellent given that they fulfill some stated criterion. Such an approach can be found in Havemann and Larsen (2015) where an excellent author is defined as (1) having a publication among the top 20% of publications that have received four or more citations per year; and (2) having co-authored an article after the first five years of his or her publication career, with twice the yearly citations of the most highly cited article during that authors' first five years. While these studies have provided valuable contributions, the question of how to define thresholds when operationalizing research excellence at the individual level with bibliometric indicators remains an open problem.

4. Previous research

This section consist of two main sub-sections. The first sub-section presents research on the predictability of research performance with bibliometric indicators, and the second sub-section consist of previous research on performance differences between males and females in science.

Predicting research performance with bibliometric indicators

This sub-section consists of a review of previous empirical research that focuses on the predictability of bibliometric indicators at the individual level. Initially some delineation of the literature under review is needed. This review is not restricted to studies focusing on the early career, since those that focus on a later career phase can also be relevant. This review focuses on studies where bibliometric indicators are used to predict future scientific output (i.e., where the output is defined in terms of some bibliometric indicator). There are some problem areas in previous research that, to some degree, overlap with the problem areas in this review, e.g., studies examining how well bibliometric indicators can predict career development (see e.g., van Dijk & Carey, 2014; Jensen, Rouquier, & Croissant, 2009), and studies examining the problem of validating peer assessments on the basis of how well the assessments can predict future values of bibliometric indicators (see e.g., Cabezas-Clavijo et al., 2013; Hornbostel et al., 2009; Nederhof, & Raan, 1987). Literature from problem areas such as the above mentioned is not included. This sub-section begins with an overview of studies that don't specifically focus on early career or doctoral education, followed by studies that do focus on early career, and finally a category of studies that examine the doctoral education.

The later career

This section presents studies that do not specifically focus on predictors in the early career, but rather include published authors in all career phases.

A relatively early study was conducted by Hirsch (2007), who examined how well the h-index, the total number of papers, the total number of citations, and the mean number of citations per paper for 50 physicists during the first 12 years of their careers could predict their performance in terms of the four above stated indicators during the subsequent 12 years. The results showed low to moderate correlations between the four indicators, with the h-index predicting itself in a future time-period and the indicator "the total number of citations" predicting itself, having the strongest correlation coefficients. Following the design of Hirsch (2007), Hönekopp and Khan (2012) examined 85 researchers in oncology and to what degree post-2000 publication success could be predicted by their previous

publication record. Hönekopp and Khan (2012) used the h-index, the number of publications, and the sum of citations as dependent variables and used the h-index, the number of publications, and citation rate of their average papers as predictors. A combination of authors' publication volume and citation rate had the strongest predictive power. Thus, a combination of traditional bibliometric indicators outperformed the h-index. To be noted is that Hönekopp and Khan only include single-authored publications in their analyses.

Danell (2011) examined to what degree we can predict the quality of the publication output of authors active in the topics Bose-Einstein condensate and episodic memory on the basis of information from the publication track record. The study included 3,685 Bose-Einstein condensate authors and 2,345 episodic memory authors. Danell (2011) conclude that the "results indicate that if we know how frequently an author's previous publication has been cited, it is easier to predict who will write the highly cited articles than it is to predict who will write, for example, median-cited articles. In other words, highly cited authors tend to write the highly cited articles, but all authors can write uncited articles" (p. 59).

Mazloumian (2012) examined how well 10 bibliometric indicators can predict future impact in a large dataset of approximately 150,000 authors from Web of Science and found that the annual citation at the time of prediction was the best predictor of future impact. However, Mazloumian concluded that, due to the low R2, existing citations do not predict citations of future work well, and should not be given significant weight in evaluating scientific potential.

Acuna et al. (2012) attempted to predict the future h-index of scientists on the basis of their past h-index, number of publications in top journals, number of distinct journals, career length, and squared number of articles, in a dataset of 3,085 neuroscientists, 57 Drosophila researchers, and 151 evolutionary scientists. The authors concluded that their prediction model consistently outperformed a model consisting of h-index alone, and that it can be used in decision-making contexts in academia. However, Penner, Pan, Petersen, Kaski, and Fortunato (2013) showed that the design in Acuna et al. (2012) was flawed since (1) cumulative, non-decreasing measures contain intrinsic auto-correlation that overestimates the predictive power of linear models, and (2) that the predictive power of these models largely depends on career age, making it problematic to include researchers of different career ages in the same model. Empirical results from Penner et al. (2013) also indicate that performance in early career phases are less predictable with bibliometric indicators than during later career phases.

Another line of research consists of a number of studies situated in the field of computer science, with a focus on comparing and improving prediction models of research performance at the individual level, often in an experimental setting

comprising machine learning and AI techniques. While most studies in this line of research focus on ECRs (i.e., early career researchers), some focus on the later career. Nezhadbiglari, Gonçalves, and Almeida (2016) introduce a novel procedure to predict the future impact of researchers called ScholarTrendLearner. Weihs and Etzioni (2017) compare five different prediction methods and use 44 features extracted from a citation graph, co-author graph, and bibliographic metadata as predictors of an author's h-index for the time period 2006-2015; they conclude that non-linear machine learning techniques outperform approaches from previous studies (e.g., Acuna et al., 2012). Ayaz, Masood, and Islam (2018) evaluated different h-index prediction models in a large dataset in the field of computer science and found that the best performing model consisted of the following predictors: the current h-index, the average citation per paper, the number of co-authors, the number of years since the first published article, publication volume, the number of journal impact factor publications, and the number of publications in distinct journals.

In summary, even if the number of studies in the later career is rather heterogeneous in terms of data, method and design, there seems to be some evidence for continuity and thus predictability, between past performance (i.e., operationalized as bibliometric indicators), and future performance. There also seems to be some evidence that a combination of indicators generates better predictions than a single indicator.

The early career

Havemann and Larsen (2015) tested 16 bibliometric indicators as predictors of later success in a sample of ECRs in astrophysics. The size of the sample was 27 later stars (a star was defined as a researcher that co-authored a highly cited paper after their first five years) and 74 random authors. Havemann and Larsen (2015) concluded that a normalized citation based indicator (i.e., an indicator that is standardized with respect to the age of the article and the expected citation behaviour in astrophysics) is best at identifying who will co-author a highly cited paper in their later career, and that this indicator can be helpful in predicting later successful authors.

In an earlier study, Symonds (2004) tested the hypothesis that publications in the high prestige journals Nature or Science, especially in the early career, increases the likelihood of producing substantial volumes of internationally regarded works in the future. The data consisted of publication track records of authors from life sciences departments in British and Australian universities in the fields of ecology and evolutionary biology. Symonds (2004) found that there is a difference in future publication volume and the propensity to publish in Nature or Science between those who had and had not published in either Nature

or Science in their early career. ECRs who published in Nature or Science were, on average, twice as productive and had over six times as many publications in Nature or Science compared to those that had not published in these journals in their early career. Symonds (2004) concludes that publications in "Nature and Science are good indicators of productive scientists, and that funding bodies and universities are justified in holding those papers, and their authors, in high regard" (p. 564). Even though Symonds' (2004) results are interesting, the study is of the document type Letter, indicating that there was no peer review involved. However, a recent study (Bornmann & Williams, 2017a) on journal prestige found similar results. Bornmann and Williams (2017a) examined how well the journal impact factor can identify ECRs who will become successful later in their careers. The results indicate that the journal impact factor can discriminate between researchers who published papers later in their careers with a citation impact above or below average, not only in the short term but also in the long term. The authors concluded that the journal impact factor should not be used as the only criterion for identifying future success, since other criteria (e.g., novelty, academic distinctions, reputation of previous institutions) should also be taken into consideration.

Lee (2019) examines the predictability of future performance (i.e., operationalized as bibliometric indicators) in a dataset consisting of 4,012 early career authors in computer and information science. The author concludes that publication volume (i.e., journal articles and conference proceedings) is the best predictor of future performance as operationalized by publication volume, as well as when future performance is operationalized by a citation-based indicator.

Another line of research focusing on the predictability of research performance in the early career revolves around social network analysis and identifying rising stars, i.e., promising ECRs in these networks. This line of research is founded within the computer sciences, rather than within the field of scientometrics. Li, Foo, Tew, & Ng (2009) developed an algorithm called PubRank to identify rising stars in social networks on the basis of co-author relationships and data from authors' publication track records (i.e., publication and citation data). Li et al. (2009) compute PubRank scores on the basis of data on early career computer scientists from 1990-1995 and compare these scores with PubRank scores in 2006. The results suggest that those authors with high PubRank scores in the first time period are more likely than the average researcher to have high PubRank scores in the later period.

Billah (2013) utilizes social network analysis to identify rising stars as defined by the h-index on the basis of a weighted co-authorship network in the field of computer science, and concludes that "the success of a young individual researcher largely depends on his or her early citations, number of collaborators,

and the impact and recent research activity of the collaborators" (p. 55). Billah and Gauch (2015) develop a classifier with the aim of identifying emerging researchers based on social network data and the h-index. The authors built an SVM classifier and evaluated it on a large set of computer science researchers in their early careers. The results showed that rising stars can be more accurately predicted based on factors related to their professional networks than factors related to their publication track records (Billah & Gauch, 2015).

In a series of publications, Daud et al. (see e.g., Daud, Abbasi, & Muhammad, 2013; Daud, Ahmad, Malik, & Che, 2015; Amjad, Daud, Khan, Abbasi, & Imran, 2018; Daud et al., 2019), examine ways to improve algorithms and methods for identifying rising stars in social networks. This line of research revolves around machine learning techniques and experimental comparisons of prediction models, and the findings suggest that, for example, mutual influence and information on publication venues might improve previous algorithms (e.g., PubRank; Li et al., 2009).

In summary, similar to the studies on the later career phases, there seems to be some evidence for continuity between performance, i.e., bibliometric indicators, during the early career and the later career. However, we should take into consideration the fact that the studies are heterogeneous in terms of data, method and design. Similar to the studies in the later career category (e.g., Hirsch, 2007; Hönekopp & Khan, 2012; Danell, 2011), Havemann and Larsen (2014) found that a citation-based indicator worked best in identifying who will produce highly cited works in the future. There seem to be some evidence that the prestige of the journal in which early career researchers publish is predictive of future performance, both in terms of publication volume and impact.

The doctoral education

In a study examining predictors of performance (i.e., publication volume) among 152 ECRs in management, Williamson and Cable (2003) concluded that the number of publications during the doctoral education has a positive effect on early career performance (i.e., publication volume). Pinheiro, Melkers, and Youtie (2014) examine the effect of publishing during doctoral education among 1,310 researchers in the United States, on research performance after attaining their PhD (i.e., the average number of publications per year after PhD), and found that publishing at least once during doctoral studies had a positive effect on future research performance. Laurance, Useche, Laurance, & Bradshaw (2013) examined the effect of gender, native language, the prestige of the institution where the researcher attained their PhD, the date of their first publication, and their pre-PhD publication record on future publication volume in a sample of 203 biologists who completed their PhDs before 2000. The researchers were affiliated

with institutions in North America, South America, Europe, and Australia. Laurance et al. (2013) concluded that: (1) out of the six predictors, early publication volume is the best predictor for future publication volume, and that counting early career publications seem to be an effective approach to identify rising stars; (2) those who publish early in their career tend to publish larger quantities in their later career; and (3) the importance of early training and mentorship for long-term publication success. Hilmer and Hilmer, (2009), and Buchmueller, Dominitz, and Lee Hansen (1999), conclude that publication volume during doctoral education has a positive effect on future publication performance in the field of economics. A common denominator of these five studies is the focus on publication volume and the conclusion that publishing during doctoral education has a positive effect on future publication volume.

Haslam and Laham (2009) examine publication records of 60 social psychologists in the USA to determine whether performance, as operationalized by publication and citation-based indices, during the doctoral education, can predict later research performance (i.e., publication and citation-based indices), and concludes that performance during doctoral education has a positive effect on future research performance. Horta and Santos (2016) examined the effect of performance (i.e., publication volume) during doctoral studies on researchers' future knowledge production and impact in a dataset comprising 664 PhDs in Portugal. Horta and Santos (2016) conclude that the number of publications during doctoral studies has a positive effect on future knowledge production and impact. Both Haslam and Laham (2009) and Horta and Santos (2016) comes to the conclusion that performance (i.e., publication volume) during doctoral studies has a positive effect on future impact as operationalized by a citation-based indicator.

In summary, similar to the studies in the category for later career phases and the category for the early career phase, the studies in the time period prior to thesis completion are few and heterogeneous in terms of data, method, and design. In line with the other two categories, there also seems to be some evidence for continuity between past and future performance, i.e., performance during the doctoral education and later career phases. Weighing together previous research on the later career phase, the earlier career phase, and the doctoral education it seems to be consensus in the results of continuity between research performance as operationalized by bibliometric indicators during doctoral education, the early career, and later in the career. A final note of caution is that many of the reviewed studies use the h-index as the main predictor and/or the outcome variable. It should be noted that the h-index has endured strong criticism as a reasonable measure of research performance at the individual level and can be problematic, depending on the design of the study, to use in prediction models (see e.g., Bornmann & Daniel, 2009; Waltman & van Eck, 2012; Penner et al., 2013).

Previous research on gender differences in research performance

Research on gender differences in research performance is positioned in a larger body of literature concerned with gender bias in science and gender differences in career development (see e.g., European Commission, 2012; Ceci, Ginther, Kahn, & Williams, 2014). This section consists of two sub-sections. The first presents a brief overview of the main issues of gender differences in career development and a theoretical framework consisting of four classes of potential explanations for these differences. The second section is more specifically concerned with previous research on gender-based performance differences.

Gender differences in career development

There is general consensus that the career prospects for females in science and academia does not equal the career prospects of males (European Commission, 2012; Ceci et al. 2014). Previous literature indicates that: (1) there are gender inequalities in education with respect to the distribution of males and females over fields of study (European Commission, 2012); (2) females are especially underrepresented in those fields that are maths intensive, e.g., geoscience, engineering, economics, mathematics/computer science, and the physical sciences (Ceci et al., 2014); (3) the proportion of females in leading positions within science and academia is generally lower than the proportion of males. This phenomenon is often referred to as the "glass ceiling" or "sticky floor" effect, which states that, for some reason, fewer females reach influential positions in science and academia than their male counterparts, e.g., senior positions, leadership positions, or become members of scientific commissions (European Commission, 2012); (4) females have, on average, lower wages than males; (5) studies have indicated that in general the proportion of females that apply for research funding is lower than the proportion of males (European Commission, 2009; European Commission, 2012); (6) female researchers tend to publish less than their male counterparts (Cole & Zuckerman, 1984; Long 1992; Xie & Shauman, 1998; Prpić, 2002; Fox, 2005).

According to the European Commission (2012), the literature lists two main concerns: (1) gender equality seems to progress slower in science and academia than in other highly skilled professions; and (2) the phenomenon of gender inequality is at odds with the norms of universalism and meritocracy that are presumably founding norms in science and academia.

Previous research has provided numerous explanations for gender inequalities in science and academia. Zuckerman (2001) divides these different explanations in four classes: (1) scientific ability; (2) social selection, which has two sub-classes, (a) gender discrimination, and (b) differences based on role performance and

differences based on how resources and rewards are distributed; (3) self-selection, which also has two sub-classes, (a) self-selection related to family, marriage, and parenthood; and (b) self-selected career commitment; (4) differences that are a consequence of cumulative advantages and disadvantages.

The first class is the scientific ability explanation, which suggests that female and male researchers have different biological and psychological characteristics, and that these differences in ability are causing the gender differences in career development in science. However, as noted by van den Besselaar and Sandström (2016), the scientific ability explanation has been contested in recent research, since a direct gender ability effect could not be found (Xie & Shauman 1998).

The second class revolves around processes of social selection. These processes involve decision making about individuals over which they have no control. This class of explanation can be contrasted with processes of self-selection, where decisions are controlled by individuals and only indirectly by "socially structured arrangements for selection" (Zuckerman, 2001, p. 71). Gender discrimination occurs in processes of social selection when females and males are treated unequal based on their gender. Discrimination can have a gender effect on career development if males' and females' opportunities for role performance is unequal, if their role performance is judged according to different standards, or if the rewards differ for the same role performance. Another form of discrimination concerns differences in how the work of female and male researchers are assessed. If female research is assessed on the basis of harsher standards than those applied for males, it may lead to less resources and slower career development (Zuckerman, 2001).

The third class revolves around explanations of gender differences based on self-selection. This class suggest that the differences in career development of males and females in science and academia can be explained by the decisions individual researchers make for themselves. The different explanations in this class focus on marriage, children, parenthood, domestic and parental roles, and how females tend to make career decisions that benefit family life but may harm their careers (Zuckerman, 2001).

The fourth class of explanation revolves around processes of cumulated advantages and disadvantages. This explanation has its basis in the theory of cumulative advantages and can be summarized as a process where social selection and self-selection factors accumulate over time and where each step, on average, provides advantages for males and disadvantages for females (Zuckerman, 2001).

Gender differences in productivity

Many empirical studies have examined performance differences between male and female researchers (see e.g., Cole & Zuckerman, 1984; Long 1992; Xie & Shauman 1998; Prpić, 2002; Fox, 2005). Even though the results have been somewhat contradictory, most have shown that all else been equal, male researchers tend to publish more papers and receive more citations than female researchers, i.e., on average males perform better than females (European Commission, 2012). Several explanations for the performance difference have been proposed; however, the results point in different directions and the general view is that gender-based performance differences have not been properly explained. The research problem of performance differences between males and females is therefore often referred to as the "the productivity puzzle" (Cole & Zuckerman, 1984).

One explanation revolves around parenthood and to what degree having children affects gender differences in productivity. First, it has been shown that females more often interrupt their careers to have children and start a family (Prozesky, 2008). Many studies have tested the relationship between the gender productivity gap and research performance based on family-related variables with mixed and somewhat contradictory results (Ceci et al., 2014). For example, Xie and Shauman (2003), Stack (2004), and Sax et al. (2002) found that having children had no effect or relatively small effects on productivity for both females and males. Another line of study suggests that having children does have a negative effect on female research performance compared with male performance (see e.g. Fox, 1995; Fuchs, von Stebut, & Allmendinger, 2001; Hunter & Leahey, 2010; Ecklund &Lincoln, 2011; Lutter & Schröder, 2019). However, Hargens, McCann, and Reskin (1978) found that children (significantly) slowed the productivity of men and women equally, and Ginther and Kahn (2006) indicate that female assistant professors in social science and geoscience, but in no other fields, who had children, publish fewer papers than women without children suggesting that there are (1) field specific differences, and (2) that having children does not have a strong influence in most fields and cannot explain the overall gender productivity gap. This line of explanations involves, as suggested by Zuckerman (2001), processes of both self-selection and social selection. Arguably individuals make their own decisions about when to marry and start a family, and how to manage their domestic and professional lives. However, these decisions may be affected by parental roles, local traditions, labour market policies, etc.

Another explanation for gender differences in productivity revolves around male and female involvement in professional networks and collaboration. It has been shown that collaboration, in terms of co-authorships, tends to increase productivity such that those who collaborate more also are co-authors in more publications (Ceci, et al. 2014). It has further been shown that females tend to

collaborate less than males (European Commission, 2012). Consequently, the lower degree of collaboration among females may explain their lower productivity.

A third explanation revolves around how time for research is distributed across males and females at different stages in their careers. Ferriman, Lubinski, and Benbow (2009), and Lubinski and Benbow (2006) compare actual and preferred work hours per week by gender in the natural sciences and engineering and found that females both preferred and actually worked fewer hours per week compared with males. Xie and Shauman (2003) more specifically investigate time for teaching versus time for research and its effect on research productivity among males and females, and found that the effect of gender on productivity decreases when time for teaching and timed for research is controlled. Similar results have been found in more recent studies. For example, Misra, Lundquist, Holmes, and Agiomavritis (2011) found that males spent almost twice as much time on research than females at a faculty of science, technology, engineering, and mathematics, and less time on teaching, advising and service (for example, attending meetings, serving on committees, emailing, organizing workshops, participating in forums). In the field of economics, Manchester and Barbezat (2013), found that male economists spend more time on research and concentrate their research more on the academic year then female economists, who concentrate their research in the summer months. Manchester and Barbezat (2013) also found that time allocation and time concentration is associated with productivity differences between males and females, where females produced fewer articles than their male colleagues.

The issue of time for research, where females tend to spend less time on research and more on other tasks, e.g., teaching, is related to another important gender issue that revolves around gender and position in the academic hierarchy access to resources that those positions enable. In fact, controlling for factors such as academic track, academic position, type of institution, and available resources, Xie and Shauman (2003), found that the observed productivity differences between males and females disappeared. Xie & Shauman (2003) conclude that gender-based differences in productivity are explained by academic position and the rewards that comes with it. If productivity differences between males and females are explained by academic position, the given question should be by what mechanisms are academic positions determined? In their study, van den Besselaar and Sandström (2016), ask if differences in career development (i.e., academic position) can be explained by previous research performance (i.e., productivity and a citation-based indicator) and conclude that this is actually the case – research performance has a positive effect on career development. Comparing Xie and Shauman's (2003) and van den Besselaar and Sandström's (2016) studies, it is interesting to notice their diametrically opposed view of the

direction of the causal relation between performance and career development. While Xie and Shauman (2003) assume that academic position explains gender differences in research productivity, van den Besselaar and Sandström (2016) assume that research productivity explains gender differences in academic position. If anything, this demonstrates the complexity of the issue at hand. In order to interpret these results, we can turn to the theory of cumulative advantages/disadvantages (Zuckerman, 2001) and Merton's resource-and-reputation model (European Commission, 2012; DiPrete & Eirich, 2006). The theory of cumulative advantage and the resource-reputation-model suggests that within the scientific system, research achievements are rewarded with access to resources, e.g., funding or academic positions. In turn, these resources enable scientific productivity and new achievements, which may provide access to even better positions and enable more resources, and so forth. The explanation of cumulative advantages/disadvantages would suggest that small initial differences between males and females cumulate over time. This logic suggests that in order to investigate the problem of productivity differences between males and females we should focus on the early stages of the career, and better yet, the phase in the career where females and males have the same academic position, but where they can vary in research performance, i.e., the doctoral education. This is a motivating factor in this thesis for examining productivity differences between males and females as early as during the doctoral education and how these differences affect performance in the early career.

The importance of doctoral education and the early career for gender differences is underlined by the European Commission (2012), where it is suggested that the early career, i.e., the years following doctoral education, can be viewed as "the rush hour" of academic careers, since small differences in the early career "may turn into wide differences in the allocation of opportunities for doing research and have a determining impact on career outcomes" (European Commission, 2012, p. 102). This line of argument is supported by Ledin et al. (2007) who found that gender-based performance differences among applicants to EMBO (i.e., European Molecular Biology Organization) fellowships can be explained by factors related to family and parenthood and the constraints these factors put on time for conducting research and career related mobility (for example, moving for a postdoc). Mason and Goulden (2004) conducted a longitudinal study on the long-term impact of family (which imply time and mobility constraints) on the careers of males and females in science, and the main conclusions were that the early career, i.e., the five years from the year of completing the thesis, is the most important time period for attaining recognition, and that females who did not have children or had children at a later stage attained tenure earlier than females who had children earlier (Mason & Goulden, 2004).

The remaining part of this section consists of an overview of the literature that focus on the early career, followed by an overview of literature focusing specifically on doctoral education. A number of studies indicate that females have lower productivity than males in their early career. In an earlier study Long (1992) examined productivity differences between males and females in biochemistry and found increasing productivity differences during the first 10 years of the publication career. A more recent study, Symonds et al. (2006), examines a sample of 168 researchers in the life sciences and shows that productivity differences appear as early as two years after early career researchers' first publication.

The effect of gender is somewhat inconclusive in previous research on the relationship between research performance during doctoral educaton and the early career. Some studies indicate that males perform better (see e.g., Hilmer & Hilmer, 2009; Laurance et al., 2013; Pinheiro et al., 2014; Horta and Santos, 2016). Other studies indicate that there is no effect, or at least not a statistically significant effect (see e.g., Buchmueller et al., 1999; Williamson & Cable, 2003; Haslam & Laham, 2009). However, the effect of gender is generally not the main focus in these studies. Rather, gender is treated as a control variable. An important contribution of this thesis is that it specifically focus on the gender variable and connects gender differences during the early career with performance differences during doctoral education and the predictability of research performance.

Finally, previous research that specifically focuses on performance differences during doctoral education is scarce. However, a recent study (Pezzoni et al., 2016), examines gender differences among six cohorts of doctoral students between 2004 and 2009 at the California Institute of Technology, with a focus on how (1) the gender of the advisor; (2) the gender pairing between the advisor and the doctoral student; and (3) the gender composition of the research team affect the performance of doctoral students. The results of Pezzoni et al. (2016) indicate that (1) female doctoral students publish, on average, fewer papers than males; (2) students with female advisors publish more; and (3) gender pairing matters: male students working with female advisors publish more than male students working with male advisors, and women students working with male advisors publish less. The authors found no indications that the gender composition of the team affected performance (i.e., productivity) during doctoral education. Pezzoni et al. (2016) also found that (1) publishing during the first year of a student's PhD studies is predictive of the student's publications in his or her subsequent graduate career, and (2) students who work with a highly productive advisor publish more than those who work with a less productive advisor. In another recent study, Feldon, Peugh, Maher, Roksa, & Tofel-Grehl (2017) concludes that despite spending more time on assigned tasks (for example, laboratory work),

females were less likely than males to publish in their first year during the doctoral education. Epstein and Lachman (2018) found that doctoral students integration into their scientific community increased male doctoral students' first authored publications, but not female doctoral students' first authored publications.

In summary, the literature on gender differences in productivity in the early career and during the doctoral education seem to largely follow the same pattern as observed in the literature concerned with the later career phases; there are gender based differences in productivity where males tend to publish more than females.

5. Results: summary of the papers

Article I

The information value of early career productivity in mathematics: a ROC analysis of prediction errors in bibliometrically informed decision making

This article treats prediction of future productivity among early career mathematicians as a binary classification problem, i.e., the prediction task is to classify researchers as members or non-members in a future top performance group.

In order to classify a researcher as a member of a top performing group it is necessary to determine a decision threshold, i.e., some minimum level of early career productivity required to be classified as a top performer. Researchers with a productivity above the decision threshold are considered top performers. However, in order optimize the choice of decision threshold we need to know the consequences of each decision threshold and choose the threshold with the best consequences according to the preferences of the decision maker.

The article investigates (1) the effect of three different definitions of top performance groups: top 10%, top 25%, and top 50%; (2) the consequences of using different thresholds in the prediction models; and (3) the added prediction value of information on early career research collaboration and publications in prestige journals.

The results show that early career productivity has an information value in all tested definitions of top performance groups. However, future performance is more predictable if the definition of a high-performance group is more exclusive, i.e., it is easier to make good predictions for the top 10% group than for the top 50% group.

A comparative analysis between the decision thresholds provided by the Youden index, which takes consequences into consideration, and a method commonly used in evaluative bibliometrics, which does not take consequences into consideration when determining decision thresholds, indicated that differences are trivial for the top 25% and 50% groups. However, a statistically significant difference between the methods was found for the top 10% group.

Information on early career collaboration and publications in prestige journals did not add any prediction value to the bibliometric indicator (i.e., publication rate) in any of the models.

The key contributions of this article are the focus on consequences in terms of prediction errors and the notion of transforming uncertainty into risk when we are choosing decision thresholds in bibliometrically informed decision making.

Article II[4]

Predicting research excellence at the individual level: The importance of publication rate, top journal publications, and top 10% publications in the case of early career mathematicians

In this article, two questions are examined: (1) whether the behaviour of publishing in top journals, in addition to publishing many articles, is an important predictor of who will attain excellence during the first eight years of their career among mathematicians, and (2) which publication behaviours (i.e., to publish a lot, to publish excellent research, to publish in high prestige journals) during the first four years of the publication career might be the most important in predicting whether an author will attain excellence in their later career.

The results for question (1) indicate the importance of publishing in top journals and that those who publish many articles in top journals, which implicitly requires a high publication count, have a higher probability of attaining excellence during the first eight years of their career. Those authors with no top journal publications had very slim chances of attaining excellence. These results suggest that publishing in top journals is particularly important in the process of attaining excellence in the early career in addition to publication volume.

The results for question (2) indicate that those authors with many publications in top journals and many top 10% publications (i.e., excellent publications), which implicitly require a high publication volume, have a higher probability of attaining future excellence than those with fewer top journal, and top 10% publications. Publication volume was redundant in all models containing predictors for top journal publications and/or top 10% publications (i.e., excellent publications), indicating that if we have information on the number of top journal publications and/or the number of top 10% publications of an author, the additional information of publication volume will not provide much help in predicting whether an author will attain future excellence.

The results also suggest that publishing excellent research and publishing in top journals is more or less equally important in terms of predicting future excellence, and that neither add much to the other in terms of predicting future excellence.

[4] An error in the data was identified after this article was published. The error did not change the results och the conclusions significantly. However, an erratum was published. This erratum is included in this thesis toghether with Article II.

Publishing in top journals and publishing excellent work seem to be interchangeable as predictors of future excellence.

Article III

Early career performance and its correlation with gender and publication output during doctoral education

Aginst the backdrop of the changing political climate in science and academia, where the publication output during doctoral education is increasingly used in selection processes for funding and employment in the early career, this article examines (1) how performance (defined as publication volume and research excellence) during the doctoral education affect the probability of attaining research excellence in the early career (i.e., the years succeeding thesis completion); and (2) if there is performance differences between males and females in the early career and to which degree these gender differences can be explained by performance differences during the doctoral education.

The dataset consisted of doctoral students employed at the faculty of science and technology and the faculty of medicine at a Swedish university. The main results indicate that:

(1) performance during the doctoral education has a positive effect on the probability of attaining research excellence in the early career;

(2) there is an interaction between publication volume and research excellence during doctoral education indicting that the effect of publication volume depends on whether a doctoral student have attained excellence or not. For the group of doctoral students who did not attain relative excellence during doctoral studies, the publication volume is a weak predictor of future excellence. However, for those who did attain excellence during doctoral studies, publication volume is a strong predictor of future excellence.

(3) males have a higher probability of producing excellent research in their early career than females; and that

(4) the differences in performance between males and females in their early career can partly be explained by performance differences during their doctoral education.

These results are interesting since they indicate that gender differences in research performance might be determined as early as during the doctoral education and that previous research has shown that even small performance

differences between males and females in the early career, tend to increase and become significant over time.

Article IV

The importance of collaboration and supervisor behaviour for gender differences in doctoral student performance and early career development

The aim of this study is to gain a better understanding of gender differences in productivity in science, technology and medicine, and how these differences arises as early as during the doctoral education. The data consisted of doctoral students that completed their education at a Swedish university between 2006 and 2010.

The article focuses on how factors related to the scientific standing of the doctoral student's supervisor and integration into collaborative networks affects performance difference between males and females during doctoral education. More specifically we (1) examine to what degree gender differences are affected by choice of database and different performance indicators; (2) examine potential explanations of the observed performance differences among female and male doctoral students; and (3) describe how the observed gender differences in publication rate is reproduced in the early career phase. The sample consists of doctoral students from the faculty of natural science and technology, and the faculty of medicine at a Swedish university.

The main results indicate that:

(1) gender difference in research productivity is evident as early as during doctoral education. The observed difference is robust with regard to choice of database, counting scheme, and weighting the publication volume with normalized citation scores of the papers;

(2) most of the observed productivity difference between males and females during doctoral education is explained by variables that express different aspects of collaborative behaviour. The variables with the largest effect sizes are: (a) the number of articles co-authored with the supervisor; (b) the number of external co-authors of the doctoral student; and (c) the collaboration coefficient (i.e., a weighted mean of the number of co-authors per paper);

(3) continued collaboration with the supervisor in the early career (i.e., the years following completion of the doctoral studies) reduces the gender difference in performance, indicating that after the completion of doctoral studies the

estimated gender gap in publication volume is also related to collaboration with the supervisor;

(4) during doctoral education and in the early career, the number of internal and external collaborators, i.e., the internal and external collaboration network, is an important explanation for performance differences between males and females.

In summary, collaborating with the supervisor and the size of the external collaborative network are the main explanations for differences in productivity during doctoral studies. For those doctoral students who continue at the same university at which they graduated, supervisor behaviour and collaboration continue to be important factors in explaining gender differences in productivity during the early career as well.

6. Concluding discussion

This thesis revolves around three overarching research questions: (1) the degree to which research performance, as operationalized by bibliometric indicators, during doctoral education and early career is predictable; (2) the degree to which gender differences in early career performance can be explained by research performance during the doctoral education; and (3) to what degree other factors, e.g., collaboration and supervisor behaviour, might affect gender differences in research performance during the doctoral education and the early career.

Concerning the first question, all four articles (Articles I, II, III, and IV) included in this thesis indicate that there is continuity between research performance, as operationalized by bibliometric indicators, during doctoral education and the early career, and performance, as operationalized by bibliometric indicators, later in the career. The main results suggest that:

(1) publication volume in the later career can be predicted by publication volume during the early career, i.e., those who publish a lot in their early career are more likely to publish more later in their career than those who publish less in their early career;

(2) future performance is more predictable if the definition of the high-performance group is more exclusive, i.e., it is easier to make good predictions if the threshold for belonging to the top group is, for example, the 10% most productive researchers, than the top 50% most productive researchers;

(3) research excellence, as operationalized by a citation-based indicator, in the later career can be predicted by bibliometric indicators during doctoral education and the early career, i.e., those who publish larger volumes, publish more in high prestige journals, and more excellent research, are more likely to attain excellence in their later career than those with lower values on these variables;

(4) if we want to identify excellent researchers, we must take both the dimension of quality, i.e., citation-based indicators, and the dimension of quantity, i.e., publication volume, into consideration. In Articles II and III, the results suggest that publication volume during doctoral education and the early career is a necessary, but not a sufficient condition for predicting future excellence. Thus, in order to identify future excellence, we should take both early publication volume and early excellence into consideration; and

(5) performance seem to be predictable in both non-collaborative disciplines where the resource dependency is low, such as mathematics (Article I and II), and

in more collaboratively driven disciplines with high resource dependence, such as physics, chemistry, engineering and the medical sciences (Article III and IV).

How do the findings of this thesis correspond with those in previous research? Previous research with a focus on researchers in the later career, the early career, and doctoral students indicates overall that there is continuity between past and future research performance as operationalized by bibliometric indicators. Weighing the results of previous research together with the results of this thesis it seems reasonable to conclude that there is continuity between research performance as operationalized by bibliometric indicators during doctoral education and the early career.

Regarding the second and third overarching questions of this thesis, i.e., the degree to which gender differences in early career performance can be explained by research performance during the doctoral education, and to what degree other factors, e.g., collaboration and supervisor behaviour, might affect gender differences in research performance during the doctoral education and the early career – which are explored in Articles III and IV – the main results indicate:

(1) that gender differences in performance can be observed as early as during doctor education and that these differences partly explain the observed gender differences in the early career; and

(2) that gender differences in performance during doctoral education can largely be explained by the doctoral student's collaborative networks and supervisor behaviour.

How do the findings related to the second and third question of this thesis line up against the findings in previous research? Previous research on this topic is scarce; however, the observed gender differences in productivity as early as during doctoral education align with the results in Pezzoni et al. (2016), Epstein and Lachmann (2018), and Feldon et al. (2017). An important contribution of this thesis is that it connects gender differences during the early career with performance differences during doctoral education and the predictability of research performance. This issue has not been much addressed in previous research on the predictability of research performance in the early stages of the research career (see e.g., Pinheiro et al., 2014; Williamson & Cable; 2003; Laurance et al., 2013; Horta & Santos; 2016), and can be considered a knowledge gap that this thesis has intended to shed some light on. Why is it important to examine this knowledge gap? Well, if gender differences in research performance as operationalized with bibliometric indicators occur as early as during the doctoral education, and these differences partly explain performance differences

in the early career, the use of these types of indicators as decision support tools might reinforce the gender gap in science.

Another important contribution of this thesis is that it provides and explanation of the observed performance differences between females and males during doctoral education as a consequence of doctoral students' collaborative networks and supervisor behaviour. An important question for future research is why factors such as publishing with the supervisor and internal and external collaborative networks affect performance differences between male and female doctoral students. To what degree are there processes of self-selection and social selection at work, and how can we make changes to enhance female performance? Since the results indicate that collaboration and collaborative networks are important factors for explaining gender differences as early as during doctoral education, an interesting question is how research collaborations come about for doctoral students, i.e., how doctoral students become co-authors on research articles, both internally, externally, and with their supervisors, what roles doctoral students have as co-authors, and if, and to what degree, these things differ between male and females.

Potential implications of research question 2 and 3 is that it may be valuable to focus on these issues in doctoral education and work towards an increased awareness among management, faculty, and supervisors of how supervisors' collaborative behaviour may affect gender differences in productivity among doctoral students. Further, it would be valuable to work towards increasing the opportunities for females to collaborate and to publish as co-authors with their supervisors, as well as increasing the opportunities for females to enlarge their external collaborative networks.

What are the practical implications of the results and conclusions of this thesis? To what degree is it reasonable to use bibliometric indicators as decision support tools in science policy and management during the early career?

The predictability of future performance based on bibliometric indicators can be said to depend on two things. First, the extent to which research performance as a concept is accurately measured by bibliometric indicators. Secondly, the extent to which future research performance (according to the definition in this thesis) depends on previous research performance.

The first point is related to the question of what bibliometric indicators measure and the quality of the data and the specific indicators (Aksnes et al., 2019). Citation theory suggests that we can interpret these indicators in terms of the concepts of impact or influence (Aksnes et al., 2019). Research performance generally refers to the concept of research quality, which is a multidimensional

concept, of which impact can be one dimension. Thus, while research performance as measured by citation-based bibliometric indicators may not measure all key dimensions of quality (as suggested by Aksnes et al., 2019) such as plausibility (i.e., the soundness and solidity of research, as well as scientific virtues such that research should be well-founded, based on scientific methods, and produce convincing results), originality (i.e., the degree to which research is novel and produces new knowledge), and societal value (i.e., the degree to which the research is relevant for, e.g., education, wealth or the environment), it seems reasonable to assume that they measure the key dimensions of scientific value, i.e., the relevance and importance of the research under assessment for future research. This aspect of what citation-based indicators measure must be taken into consideration when interpreting the results of this thesis.

The second point, i.e., the extent to which future performance depends on previous research performance, is related to a discussion on causality and explanatory mechanisms. Generally, the results of the studies in this thesis indicate that it is possible to identify who will perform well in their early career by looking at performance earlier in their career. However, the underlying mechanisms behind these relationships have not been studied in this thesis. What can be concluded is that the results in Articles III and IV indicate that when gender, a number of variables associated with supervisor behaviour and collaboration, and a number of other variables are controlled for, performance during doctoral education is still indicative of who will perform well during their early career.

The question of underlying mechanisms is especially important when bibliometric indicators are used in decision-making contexts in science policy and management, where the decision support tools provide incitements for behavioural change among those who are targeted by them. Consider the theory of cumulative advantages; if the positive relationship between performance during the doctoral education and the early career is a consequence of unwanted Matthew effects or some other bias, it would be problematic to reinforce it with policy. If, on the other hand, the cumulative advantages reinforce initial and real differences in scientific ability, then it might be warranted to actually use policy to reinforce them (Danell, 2011). To conclude, while the results of this thesis suggest that research performance during doctoral education and the early career is to some degree predictable with bibliometric indicators, there are observed gender differences in performance that have to be taken into consideration. If they are not, the use of these types of performance indicators in science policy and management might even increase the gender gap in science.

References

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2011). National research assessment exercises: a comparison of peer review and bibliometrics rankings. *Scientometrics, 89*(3), 929–941.

Abramo, G., D'Angelo, C. A., & Reale, E. (2019). Peer review versus bibliometrics: Which method better predicts the scholarly impact of publications? *Scientometrics, 121*(1), 537–554. https://doi.org/10.1007/s11192-019-03184-y

Acuna, D. E., Allesina, S., & Kording, K. P. (2012). Predicting scientific success: Daniel E. Acuna, Stefano Allesina and Konrad P. Kording present a formula to estimate the future h-index of life scientists. *Nature, 489*(7415), 201–202. https://doi.org/10.1038/489201a

Addis, E., & Pagnini, C. (2010). Meta-analysis of gender and science research. Topic Report Gender and Scientific Excellence. EGERA. (2016). Deliverable number D, 6.

Aksnes, D. W. 2003. Characteristics of highly cited papers. *Research Evaluation, 12*(3), pp.159–170.

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. *SAGE Open, 9*(1), 17. https://doi.org/10.1177/2158244019829575

Aksnes, D. W., & Taxt, R. E. (2004). Peer reviews and bibliometric indicators: A comparative study at Norwegian University. *Research Evaluation, 13*(1), 33–41.

Allen, L., Jones, C., Dolby, K., Lynn, D., & Walport, M. (2009). Looking for landmarks: The role of expert review and bibliometric analysis in evaluating scientific publication outputs. *PLoS ONE, 4*(6), e5910.

Allison, P. D., & Stewart, J. A. (1974). Productivity Differences Among Scientists: Evidence for Accumulative Advantage. American Sociological *Review, 39*(4), 596–606. https://doi.org/10.2307/2094424

Amjad, T., Daud, A., Khan, S., Abbasi, R. A., & Imran, F. (2018). Prediction of Rising Stars from Pakistani Research Communities. In 2018 *14th International Conference on Emerging Technologies (ICET)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICET.2018.8603661

Amsterdamska, O., & Leydesdorff, L. (1989). Citations: Indicators of significance? *Scientometrics, 15*(5-6), 449–471.

Australian Research Council. (2007). Early Career Researcher. http://www.arc.gov.au/applicants/researcher_early.htm

Bazeley, P. (2003). Defining 'early career' in research. *Higher Education, 45*(3), 257–279.

Bertocchi, G., Gambardella, A., Jappelli, T., Nappi, C. A., & Peracchi, F. (2015). Bibliometric evaluation vs. informed peer review: Evidence from Italy. *Research Policy, 44*(2), 451–466. https://doi.org/10.1016/j.respol.2014.08.004

Billah, S. (2013). Identifying Emerging Researchers using Social Network Analysis. ProQuest Dissertations Publishing. Retrieved from http://search.proquest.com/docview/1490797091/

Billah, S. M., & Gauch, S. (2015, November). Social network analysis for predicting emerging researchers. In 2015 *7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)* (Vol. 1, pp. 27-35). IEEE.

Bornmann, L. (2014). How are excellent (highly cited) papers defined in bibliometrics? A quantitative analysis of the literature. *Research Evaluation, 23*(2), 166-173.

Bornmann, L., & Daniel, H. (2009). The state of h index research. *EMBO Reports, 10*(1), 2–6. https://doi.org/10.1038/embor.2008.233

Bornmann, L., & Leydesdorff, L. (2018). Count highly-cited papers instead of papers with h citations: use normalized citation counts and compare “like with like”! *Scientometrics, 115*(2), 1119–1123. https://doi.org/10.1007/s11192-018-2682-1

Bornmann, L., & Leydesdorff, L. (2013). The validation of (advanced) bibliometric indicators through peer assessments: A comparative study using data from InCites and F1000. *Journal of Informetrics, 7*(2), 286–291.

Bornmann, L., & Williams, R. (2017a). Can the journal impact factor be used as a criterion for the selection of junior researchers? A large-scale empirical study based on ResearcherID data. *Journal of Informetrics, 11*(3), 788–799. https://doi.org/10.1016/j.joi.2017.06.001

Bornmann, & Williams. (2017b). Use of the journal impact factor as a criterion for the selection of junior researchers: A rejoinder on a comment by Peters (2017). *Journal of Informetrics, 11*(4), 945–947. https://doi.org/10.1016/j.joi.2017.08.005

Buchmueller, Dominitz, & Lee Hansen. (1999). Graduate training and the early career productivity of Ph.D. economists. *Economics of Education Review, 18*(1), 65–77. https://doi.org/10.1016/S0272-7757(98)00019-3

Cabezas-Clavijo, Á., Robinson-García, N., Escabias, M., & Jiménez-Contreras, E. (2013). Reviewers' ratings and bibliometric indicators: Hand in hand when assessing over research proposals? *PLoS ONE, 8*(6), e68258.

Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in Academic Science: A Changing Landscape. *Psychological Science in the Public Interest, 15*(3), 75–141. https://doi.org/10.1177/1529100614541236

Cole, S., Cole, J. R., & Simon, G. A. (1981). Chance and consensus in peer-review. *Science, 214*(4523), 881–886. doi: 10.1126/science.7302566

Cole, J. R., and Cole, S. (1973). *Social Stratification in Science*. Chicago: University of Chicago Press.

Cole, J. R., & Zuckerman, H. (1984). 'The productivity puzzle: Persistence and change in patterns of publication of men and women scientists', in P. Maehr, & M.W. Steinkamp, eds. Advances in motivation and achievement, JAI Press, Conn.

Colliander, C. (2014). *Science mapping and research evaluation: a novel methodology for creating normalized citation indicators and estimating their stability*. Umeå universitet, Umeå.

Cozzens, S. (1989). What do citations count? the rhetoric-first model. *Scientometrics, 15*(5), 437–447.

Dalton, G. W., Thompson, P. H., & Price, R. L. (1977). The four stages of professional careers— A new look at performance by professionals. Organizational *Dynamics, 6*(1), 19–42. https://doi.org/10.1016/0090-2616(77)90033-X

Danell, R. (2011). Can the Quality of Scientific Work Be Predicted Using Information on the Author's Track Record? *Journal of the American Society for Information Science and Technology, 62*(1), 50–60. https://doi.org/10.1002/asi.21454

Daud, A., Abbasi, R., & Muhammad, F. (2013). *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 7825, pp. 13–24). https://doi.org/10.1007/978-3-642-37487-6_4

Daud, A., Ahmad, M., Malik, M., & Che, S. (2015). Using machine learning techniques for rising star prediction in co-author network. *Scientometrics, 102*(2), pp. 1687–1711. https://doi.org/10.1007/s11192-014-1455-8

Daud, A., Aljohani, N. R. H., Alyoubi, K. H. A., Abbasi, R. A., Rafique, Z., Amjad, T., & Dawood, H. (2019). Finding rising stars in co-author networks via weighted mutual influence. In *26th International World Wide Web Conference 2017, WWW 2017 Companion* (pp. 32–41). International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3041021.3054137

De Bellis, N. (2009). *Bibliometrics and citation analysis: from the Science citation index to cybermetrics*. Lanham, Md.: Lanham, Md.: Scarecrow Press.

DiPrete, T.A., & Eirich, G.M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology, 32*, 271–297.

Ecklund, E., & Lincoln, A. E. (2011). Scientists want more children. *PLoS ONE, 6*(8), e22590.

Epstein, N., & Lachmann, D. (2018). Gender Inequity during the Ph.D.: Females in the Life Sciences Benefit Less from Their Integration into the Scientific Community. *Social Sciences, 7*(8). https://doi.org/10.3390/socsci7080140

Commission. (2012). *Meta-analysis of gender and science: Synthesis report*. Publication Office of the European Union: Luxembourg.

European Commission. (2009). *The Gender Challenge in Research Funding. Assessing the European national scenes*. Office for Official Publications of the European Communities: Luxembourg.

Feldon, D. F., Peugh, J., Maher, M. A., Roksa, J., & Tofel-Grehl, C. (2017). Time-to-Credit Gender Inequities of First-Year PhD Students in the Biological Sciences. *CBE - Life Sciences Education, 16*(1). https://doi.org/10.1187/cbe.16-08-0237

Ferriman, K., Lubinski, D., & Benbow, C. P. (2009). Work preferences, life values, and personal views of top math/science graduate students and the profoundly gifted: Developmental changes and gender differences during emerging

adulthood and parenthood. *Journal of Personality and Social Psychology, 97*(3), 517–532.

Fox, M. F. (2005). Gender, Family Characteristics, and Publication Productivity among Scientists. *Social Studies of Science, 35*(1), 131–150.

Fox, M. F. (1983). Publication Productivity among Scientists: A Critical Review. *Social Studies of Science, 13*(2), 285–305. https://doi.org/10.1177/030631283013002005

Fox, M. F. (1995). Women in scientific careers. In S. Jasanoff, G. E. Markle, J. C. Petersen, & T. Pinch (Eds.), Handbook of science and technology studies (pp. 205–223). Thousand Oaks, CA: Sage.

Fuchs, S., von Stebut, J., & Allmendinger, J. (2001). Gender, science, and scientific organizations in Germany. *Minerva, 39*(2), 175–201.

Garfield, E., & Welljams-Dorof, A. (1992). Of Nobel class: A citation perspective on high impact research authors. *Theoretical Medicine, 13*(2), 117-135. doi:10.1007/bf02163625

Gingras, Y., & Wallace, M. L. (2010). Why it has become more difficult to predict Nobel Prize winners: A bibliometric analysis of nominees and winners of the chemistry and physics prizes (1901-2007). *Scientometrics, 82*(2), 401–412. doi:10.1007/s11192-009-0035-9

Ginther, D. K., & Kahn, S. (2006). Women's careers in academic social science: Progress, pitfalls, and plateaus. In A. Lanteri & J. Vromen (Eds.), The economics of economists. Cambridge, England: Cambridge University Press.

Hakkarainen, K., Hytönen, K., Makkonen, J., & Lehtinen, E. (2016). Extending collective practices of doctoral education from natural to educational sciences. *Studies in Higher Education, 41*(1), 63–78. https://doi.org/10.1080/03075079.2014.914910

Hallonsten, & Hugander. (2014). Supporting 'future research leaders' in Sweden: Institutional isomorphism and inadvertent funding agglomeration. *Research Evaluation, 23*(3), 249–260. https://doi.org/10.1093/reseval/rvu009

Hammarfelt, B., & de Rijcke, S. (2015). Accountability in context: effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University. *Research Evaluation, 24*(1), 63–77. https://doi.org/10.1093/reseval/rvu029

Hammarfelt, B., Nelhans, G., Eklund, P., & Åström, F. (2016). The heterogeneous landscape of bibliometric indicators: Evaluating models for allocating resources at Swedish universities. Research *Evaluation, 25*(3), 292–305. https://doi.org/10.1093/reseval/rvv040

Hargens, L., McCann, J., & Reskin, B. (1978). Productivity and reproductivity: fertility and professional achievement among research scientists. *Social Forces, 57*(1), 154–163. https://doi.org/10.2307/2577631

Haslam, N., & Laham, S. M. (2009). Ten Years on: Does Graduate Student Promise Predict Later Scientific Achievement? *Current Research in Social Psychology, 14*(10), 143–149.

Havemann, F., & Larsen, B. (2015). Bibliometric indicators of young authors in astrophysics: Can later stars be predicted? *Scientometrics, 102*(2), 1413–1434. https://doi.org/10.1007/s11192-014-1476-3

Hicks, D. (2012) 'Performance-based University Research Funding Systems', *Research Policy, 41*(2), 251–61.

Hilmer, Michael J., & Hilmer, Christiana E. (2009). Fishes, ponds, and productivity: Student- advisor matching and early career publishing success for economics PhDs. *Economic Inquiry, 47*(2), 290–303. https://doi.org/10.1111/j.1465-7295.2007.00108.x

Hirsch, J. E. (2007). Does the h index have predictive power? *Proceedings of the National Academy of Sciences, 104*(49), 19193–19198. https://doi.org/10.1073/pnas.0707962104

Hornbostel, S., Böhmer, S., Klingsporn, B., Neufeld, J., & Ins, M. (2009). Funding of young scientist and scientific excellence. *Scientometrics, 79*(1), 171–190. https://doi.org/10.1007/s11192-009-0411-5

Horta, H, & Santos, J. M. (2016). The Impact of Publishing during PhD Studies on Career Research Publication, Visibility, and Collaborations. *Research in Higher Education, 57*(1), 28–50. https://doi.org/10.1007/s11162-015-9380-0

Huber, J. C. (2001). A new method for analyzing scientific productivity. Journal of the American Society for Information Science and Technology, 52(13), 1089–1099. doi: 10.1002/asi.1173

Huber, J. C. (1998). Cumulative advantage and success-breeds-success: The value of time pattern analysis. *Journal of the American Society for Information*

Science, 49(5), 471–476. doi: 10.1002/(SICI)1097-4571(19980415)49:5<471: AID-ASI8>3.0.CO;2-T

Huber, J. C., & Wagner-Döbler, R. (2001a). Scientific production: A statistical analysis of authors in physics, 1800-1900. *Scientometrics, 50*(3), 437–453. doi: 10.1023/a:1010558714879

Huber, J. C., & Wagner-Döbler, W. (2001b). Scientific production: A statistical analysis of authors in mathematical logic. *Scientometrics, 50*(2), 323–337. doi: 10.1023/a:1010581925357

Hunter, L. A., & Leahey, E. (2010). Parenting and research productivity: New evidence and methods. *Social Studies of Science, 40*(3), 433–451.

Hönekopp, J., & Khan, J. (2012). Future publication success in science is better predicted by traditional measures than by the h index. *Scientometrics, 90*(3), 843–853. https://doi.org/10.1007/s11192-011-0551-2

Jensen, P., Rouquier, J.-B., & Croissant, Y. (2009). Testing bibliometric indicators by their prediction of scientists' promotions. *Scientometrics, 78*(3), 467–479. https://doi.org/10.1007/s11192-007-2014-3

Kettle, D.F. (2005). *The Global Public Management Revolution*. Brookings Institution Press.

Kärki, R., Kortelainen, T., Eriksson, K. & Ungern-Sternberg, S. (1998). *Introduktion till bibliometri*. Helsingfors: NORDINFO.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, Mass.: Harvard Univ. Press.

Laudel, G, & Gläser, J. (2008). From Apprentice to Colleague: The Metamorphosis of Early Career Researchers. *Higher Education: The International Journal of Higher Education and Educational Planning, 55*(3), 387–406. https://doi.org/10.1007/s10734-007-9063-7

Laurance, W., Useche, C., Laurance, S., & Bradshaw, C. (2013). Predicting Publication Success for Biologists. *BioScience, 63*(10), 817–823. https://doi.org/10.1525/bio.2013.63.10.9

Ledin, A., Bornmann L., Gannon F. & Wallon G. (2007) A persistent problem. Traditional gender roles hold back female scientists. *EMBO reports, 8*(11), 982–987.

Lee, D. H. H. (2019). Predicting the research performance of early career scientists. *Scientometrics, 121*(3), 1481–1504. https://doi.org/10.1007/s11192-019-03232-7

Li, X. L., Foo, C. S., Tew, K. L., & Ng, S. K. (2009). Searching for rising stars in bibliography networks. *In International conference on database systems for advanced applications* (pp. 288-292). Springer, Berlin, Heidelberg.

Long, J. (1992). Measures of sex difference in scientific productivity. *Social Forces, 71*(1), 159–178. https://doi.org/10.2307/2579971

Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for math–science expertise. Perspectives on Psychological Science, 1, 316–345.

Lutter, M., & Schröder, M. (2019). Is there a motherhood penalty in academia? The gendered effect of children on academic publications. IDEAS Working Paper Series from RePEc. Retrieved from http://search.proquest.com/docview/2186320808/

Luukkonen, T. (1990). *Citations in the rhetorical, reward, and communication systems of science* (Doctoral thesis). University of Tampere, Finland.

Lärarhögskolan. (2019). Medel för forskningstid inom det utbildningsvetenskapliga området. Umeå: Umeå Universitet.

Manchester, C., & Barbezat, D. (2013). The Effect of Time Use in Explaining Male–Female Productivity Differences Among Economists. *Industrial Relations: A Journal of Economy and Society, 52*(1), 53–77. https://doi.org/10.1111/irel.12011

Martin, B. R., & Irvine, J. (1983). Assessing basic research: Some partial indicators of scientific progress in radio astronomy. *Research Policy, 12*(2), 61–90.

Mason, S. (2018). Publications in the doctoral thesis: challenges for doctoral candidates, supervisors, examiners and administrators. *Higher Education Research & Development, 37*(6), 1231–1244. https://doi.org/10.1080/07294360.2018.1462307

Mason, M. A., & Goulden, M. (2004). Do babies matter (Part II)? Closing the baby gap. *Academe, 90*(6), 10–15.

Mazloumian, A. (2012). Predicting Scholars' Scientific Impact. *PLOS ONE, 7*(11): e49246. https://doi.org/10.1371/journal.pone.0049246

McGrail, M. R., Rickard, C. M., & Jones, R. (2006). Publish or perish: A systematic review of interventions to increase academic publication rates. *Higher Education Research & Development, 25*(1), 19–35.

Merton, R. K. (1968). The Matthew Effect in Science. *Science, 159*(3810), 56–63. https://doi.org/10.1126/science.159.3810.56

Merton, R. K. (1988). The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property. *Isis, 79*(4), 606–623. https://doi.org/10.1086/354848

Merton R. K. (1973). The normative structure of science. In The Sociology of Science, ed. NW Storer, pp. 267–78. Chicago: Univ. Chicago Press.

Misra, J., Lundquist, J., Holmes, E. D., & Agiomavritis, S. (2011). Associate professors and gendered barriers to advancement. *Academe, 97*(1).

Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht: Dordrecht: Springer.

Nederhof, A. J. (1988). The validity and reliability of evaluation of scholarly performance. In A. F. J. van Raan (Ed.), Handbook of Quantitative Studies of Science and Technology (pp. 193–228). Amsterdam: Elsevier.

Nederhof, A., & van Raan, A. F. J. (1987). Peer review and bibliometric indicators of scientific performance: A comparison of cum laude doctorates with ordinary doctorates in physics. *Scientometrics, 11*(5), 333–350. https://doi.org/10.1007/BF02279353

Nezhadbiglari, M., Gonçalves, M., & Almeida, J. (2016). Early Prediction of Scholar Popularity. In Proceedings of the 16th ACM/IEEE-CS on joint conference on digital libraries (Vol. 2016, pp. 181–190). ACM. https://doi.org/10.1145/2910896.2910905

Oppenheim, C. (1997). The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology. *Journal of Documentation, 53*(5), 477–487. https://doi.org/10.1108/EUM0000000007207

Oppenheim, C., & Norris, M. (2003). Citation counts and the research assessment exercise V: Archaeology and the 2001 RAE. *Journal of Documentation, 56*(6), 709–730.

Penner, O., Pan, R., Petersen, A., Kaski, K., & Fortunato, S. (2013). On the Predictability of Future Impact in Science. *Scientific Reports, 3*(1), 3052. https://doi.org/10.1038/srep03052

Petersen, A., Jung, W., Yang, J., & Stanley, H. (2011). Quantitative and empirical demonstration of the Matthew effect in a study of career longevity. *Proceedings of the National Academy of Sciences, 108*(1), 18–23.

Petersen, A., Riccaboni, M., Stanley, H., & Pammolli, F. (2012). Persistence and uncertainty in the academic career. *Proceedings of the National Academy of Sciences, 109*(14), 5213–5218.

Pezzoni, M., Mairesse, J., Stephan, P., & Lane, J. (2016). Gender and the Publication Output of Graduate Students: A Case Study. *PLoS ONE, 11*(1), e0145146. https://doi.org/10.1371/journal.pone.0145146

Pinheiro, Melkers, & Youtie. (2014). Learning to play the game: Student publishing as an indicator of future scholarly success. *Technological Forecasting & Social Change, 81*, 56–66. https://doi.org/10.1016/j.techfore.2012.09.008

Price, D. D. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science, 27*(5), 292–306.

Pride, D., & Knoth, P. (2018). Peer review and citation data in predicting university rankings, a large-scale analysis. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 11057, pp. 195–207). Springer Verlag. https://doi.org/10.1007/978-3-030-00066-0_17

Prozesky, H. (2008). A career-history analysis of gender differences in publication productivity among South African academics. *Science Studies, 21*(2), 47–67.

Prpić, K. (2002). Gender and productivity differentials in science. *Scientometrics, 55*(1), 27–58. https://doi.org/10.1023/A:1016046819457

Reale, E., Barbara, A., & Costantini, A. (2007). Peer review for the evaluation of academic research: Lessons from the Italian experience. *Research Evaluation, 16*(3), 216–228.

Research Assessment Exercise. (2005). Guidance on Submissions. RAE 03/2005. http://www.rae.ac.uk/pubs/2005/03/

Rinia, E., van Leeuwen, T., van Vuren, H., & van Raan, A. F. J. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria: Evaluation of condensed matter physics in the Netherlands. *Research Policy, 27*(1), 95–107. https://doi.org/10.1016/S0048-7333(98)00026-2

Riviera, E. (2013). Scientific communities as autopoietic systems: The reproductive function of citations. *Journal of the American Society for Information Science and Technology, 64*(7), 1442–1453.

Rodgers, R., & Rodgers, N. (1999). The Sacred Spark of Academic Research. Journal of Public Administration Research and Theory: *J-PART, 9*(3), 473–492. https://doi.org/10.1093/oxfordjournals.jpart.a024419

Rossiter, M. W. (1993). The Matthew Matilda Effect in Science. Social Studies of *Science, 23*(2), 325–341. https://doi.org/10.1177/030631293023002004

Rothwell, P. M., & Martyn, C. N. (2000). Reproducibility of peer review in clinical neuroscience: Is agreement between reviewers any greater than would be expected by chance alone? *Brain, 123*(9), 1964–1969.

Sax, L., Hagedorn, J., Arredondo, S., & Dicrisi, M. (2002). Faculty Research Productivity: Exploring the Role of Gender and Family-Related Factors. *Research in Higher Education, 43*(4), 423–446. https://doi.org/10.1023/A:1015575616285

Simonton, D. K. (1985). Quality, quantity, and age – the careers of ten distinguished psychologists. *International Journal of Aging & Human Development, 21*(4), 241–254. doi: 10.2190/KB7E-A45M-X8X7-DQJ4

Simonton, D. K. (1997). Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review, 104*(1), 66–89. doi: 10.1037/0033-295x.104.1.66

Simonton, D. K. (2003). Scientific creativity as constrained Stochastic behavior the integration of product, person, and process perspectives. *Psychological Bulletin, 129*(4), 475–494. doi: 10.1037/0033-2909.129.4.475

Simonton, D. K. (2004). *Creativity in science: Chance, logic, genius, and zeitgeist*. Cambridge, NY: Cambridge University Press.

Sinclair, J., Barnacle, R., & Cuthbert, D.. (2014). How the Doctorate Contributes to the Formation of Active Researchers: What the Research Tells Us. *Studies in Higher Education, 39*(10), 1972–1986. https://doi.org/10.1080/03075079.2013.806460

Sivertsen, G. (2017). Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. *Palgrave Communications, 3*(1), 1–6. https://doi.org/10.1057/palcomms.2017.78

Small, H. (2004). On the shoulders of Robert Merton: towards a normative theory of citation. *Scientometrics, 60*(1), 71–79.

Stack, S. (2004). Gender, children and research productivity. *Research in Higher Education, 45*(8), 891–920. doi:10.1007/s11162-004-5953-z

Symonds, M. R. (2004). Nature and Science know best. *Trends in Ecology & Evolution, 19*(11), 564–564. https://doi.org/10.1016/j.tree.2004.09.007

Symonds, M. R. E., Gemmell, N. J., Braisher, T. L., Gorringe, K. L., Elgar, M. A., & Tregenza, T. (2006). Gender Differences in Publication Output: Towards an Unbiased Metric of Research Performance (Gender and Publication Output). *PLoS ONE, 1*(1), e127. https://doi.org/10.1371/journal.pone.0000127

Thomas, P. R., & Watkins, D. S. (1998). Institutional research rankings via bibliometric analysis and direct peer-review: A comparative case study with policy implications. *Scientometrics, 41*(3), 335–355.

Tijssen, R. J. W. (2003). Scoreboards of research excellence. *Research Evaluation, 12*(2), 91–103. https://doi.org/10.3152/147154403781776690

Wagner, C. S., Horlings, E., Whetsell, T. A., Mattsson, P., & Nordqvist, K. (2015). Do Nobel laureates create prize-winning networks? An analysis of collaborative research in physiology or medicine. *PLoS ONE, 10*(7), e0134164. doi:10.1371/journal.pone.0134164

Waltman, L., & van Eck, N. J. (2012). The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology, 63*(2), 406–415.

van den Besselaar, P., & Sandström, U. (2016). Gender differences in research performance and its impact on careers: a longitudinal case study. *Scientometrics, 106*(1), 143–162. https://doi.org/10.1007/s11192-015-1775-3

van Dijk, Manor, & Carey. (2014). Publication metrics and success on the academic job market. *Current Biology, 24*(11), R516–R517. https://doi.org/10.1016/j.cub.2014.04.039

van Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics, 67*(3), 491–502. doi:10.1556/Scient.67.2006.3.10

van Raan, A. F. J. (1998). In matters of quantitative studies of science the fault of theorists is offering too little and asking too much. *Scientometrics, 43(*1), pp.129–139.

Weihs, L., & Etzioni, O. (2017). Learning to Predict Citation-Based Impact Measures. In 2017 ACM/*IEEE Joint Conference on Digital Libraries (JCDL)* (pp. 1–10). IEEE. https://doi.org/10.1109/JCDL.2017.7991559

Whitley, R. (1984). *The intellectual and social organization of the sciences* (1. ed.). Oxford: Oxford University Press.

Whitley, R., & Gläser, J. (Ed.). (2007). *The changing governance of the sciences: the advent of research evaluation systems*. Dordrecht: Springer.

Williamson, I. O., & Cable, D. M. (2003). Predicting early career research productivity: the case of management faculty. *Journal of Organizational Behavior, 24*(1), 25–44. https://doi.org/10.1002/job.178

Wouters, P. (1999). Beyond the holy grail: From citation theory to indicator theories. *Scientometrics, 44*(3), 561–580.

Xie, Y., & Shauman, K. A. (1998). Sex differences in research productivity: New evidence about an old puzzle. *American Sociological Review, 63*(6), 847–870.

Zuckerman, H. (1987). Citation analysis and the complex problem of intellectual influence. *Scientometrics, 12*(5), 329–338. https://doi.org/10.1007/BF02016675

Zuckerman, H. (2001). The careers of men and women scientists: Gender differences in career attainment. In: M. Wyer (Ed.) Women, science and technology: A reader in feminist science studies. Routledge: 69–78.

Acknowledgements

First and foremost, I would like to thank my main supervisor Rickard Danell for generous and excellent support throughout the whole process of writing this thesis. Your knowledge and guidance have been much appreciated.

I would also like to thank Cristian Colliander, my co-supervisor, for being generous with time, a thorough reader, and helpful with programming and other technical stuff. Thanks Cristian.

I would like thank Olle Persson who was my co-supervisor a short period in the beginning of my time as a doctoral student.

Thanks to Fredrik Åström and Björn Hammarfelt for valuable comments during the mid- and end-seminars, as well as Mikael Hjerm and Lena Karlsson for constructive feedback on the final draft.

I would like to acknowledge everybody at the Department of Sociology for providing a nice working environment, and especially everybody at Inforsk and LIS.

Thanks to my brother Andreas Stenling for encouraging me to pursue a PhD, and his family, Cissi, Bertil, and Aldor.

Jag vill tacka min mamma och pappa. Ni har under hela mitt liv varit mycket omtänksamma, stöttande, och goda förebilder. Tack! Ni har varit, och är, de bästa föräldrar en son kan önska.

Finally, I want to thank my lovely wife Malin and our incredible children, Vilja, Herman, and Ingeborg. I am so grateful for the privilege to spend my life with you.

Umeå, 29 january, 2020,

Jonas